# Analysis of components of the mitochondrial transcription machinery in *Arabidopsis thaliana*

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.) im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Diplom-Biologin Kristina Kühn
geboren am 27.08.1974 in Stollberg

Präsident der Humboldt-Universität zu Berlin in Vertretung
Prof. Dr. Hans Jürgen Prömel

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Thomas Buckhout, PhD

Gutachter: 1. Prof. Dr. Thomas Börner
2. Prof. Dr. Wolfgang Hess
3. Prof. Dr. Frank Kempken

Datum der mündlichen Prüfung: 24. Februar 2006

**Abstract**

In der vorliegenden Arbeit wurde die Transkription mitochondrialer Gene durch die kernkodierten Phagentyp-RNA-Polymerasen RpoTm und RpoTmp der Pflanze Arabidopsis untersucht.

Im Mitochondriengenom von Arabidopsis wurden für 12 Gene Promotoren bestimmt. Diese zeigten verschiedene Sequenzelemente und wichen meist von der für Dikotyle publizierten Konsensussequenz ab. Für die Mehrheit der Gene wurden multiple Promotoren identifiziert. Es wurden weiterhin Promotoren nachgewiesen, welche die Transkription vermutlich nicht funktioneller Sequenzen aktivieren. Architektur, Lokalisation und Nutzung mitochondrialer Promotoren implizieren eine wenig stringente Kontrolle der Transkriptionsinitiation in Arabidopsis-Mitochondrien.

Zur Analyse der Funktionen von RpoTm und RpoTmp wurde ein in vitro-Transkriptions-system entwickelt. Da RpoT-Enzyme möglicherweise Kofaktoren benötigen, wurde in Arabidopsis nach Genen potentieller mitochondrialer Transkriptionsfaktoren gesucht. Als mitochondriales Protein mit Ähnlichkeit zu mtTFB, einem essentiellen Transkriptionsfaktor in Hefemitochondrien, wurde MetA identifiziert. In in vitro-Assays initiierte RpoTm an verschiedenen Promotoren die Transkription, während RpoTmp keine signifikante Promotorspezifität zeigte. Die spezifische Promotornutzung durch RpoTm erforderte superhelikale DNA. Weder RpoTm noch RpoTmp wurde durch MetA stimuliert. Eine mtTFB-ähnliche Funktion von MetA ist daher unwahrscheinlich. Für MetA wurde ausserdem eine engere phylogenetische Beziehung zu nukleären rRNA-Dimethylasen als zu mtTFB ermittelt.

Die hier vorgestellten Studien belegen die Transkription mitochondrialer Gene in Arabidopsis durch RpoTm; für RpoTmp ist eine nicht-redundante Transkriptionsfunktion denkbar. Die Kofaktor-unabhängige Spezifität von RpoTm für verschiedene Promotoren und die wenig stringente Initiationskontrolle in vivo legen nahe, dass eine individuelle Regulation mitochondrialer Gene in Arabidopsis auf Transkriptionsebene nicht erfolgt.

**Schlagworte:**

Pflanzenmitochondrien

Mitochondriengenom

Promotor

Transkription

Phagentyp-RNA-Polymerasen

**Abstract**

Mitochondria depend on a nucleus-encoded transcription machinery to express their genome. The present study examined the transcription of mitochondrial genes by two nucleus-encoded phage-type RNA polymerases, RpoTm and RpoTmp, in the plant Arabidopsis.

For selected mitochondrial genes in Arabidopsis, transcription initiation sites were determined. Most genes were found to possess multiple promoters. The identified promoters displayed diverse sequence elements and mostly deviated from a nonanucleotide consensus derived previously for dicot mitochondrial promoters. Several promoters were detected that activate transcription of presumably non-functional sequences. Promoter architecture, distribution and utilization suggest a non-stringent control of transcription initiation in Arabidopsis mitochondria.

An in vitro transcription system was set up to elucidate the roles of RpoTm and RpoTmp. Since RpoT enzymes possibly require auxiliary factors, the Arabidopsis genome was screened for potential cofactors of phage-type RNA polymerases. A mitochondrial protein (MetA) with similarity to mtTFB, an essential transcription factor in yeast mitochondria, was identified. In in vitro transcription studies, RpoTm recognized various promoters whereas RpoTmp displayed no significant promoter specificity. Promoter recognition by RpoTm depended on supercoiled DNA templates. Transcription initiation by RpoTm or RpoTmp was not affected by MetA, indicating that MetA is not functionally equivalent to mtTFB. Besides, MetA was found to be more closely related to non-mitochondrial rRNA dimethylases than to mtTFB.

The present study establishes RpoTm to transcribe mitochondrial genes; RpoTmp may have a non-overlapping transcriptional role in mitochondria. The cofactor-independent promoter specificity of RpoTm and the apparently non-stringent control of transcription initiation in vivo imply that mitochondrial genes in Arabidopsis may not be regulated individually at the transcriptional level.

**Keywords:**

Plant mitochondria

Mitochondrial genome

Promoter

Transcription

Phage-type RNA polymerase

**Table of Contents**

## ZUSAMMENFASSUNG

Im Zellkern kodierte RNA-Polymerasen sind für die Transkription der mitochondrialen Genome eukaryotischer Zellen verantwortlich und übernehmen somit eine zentrale Rolle in der mitochondrialen Genexpression. Für verschiedenen Pflanzenspezies wurden nukleäre T7-phagenähnliche RNA-Polymerasegene (*RpoT*-Gene) beschrieben, welche vermutlich für katalytische Untereinheiten der mitochondrialen Transkriptionsmaschinerie kodieren. Eine mitochondriale Transkriptionsfunktion von *RpoT*-Genprodukten in photosynthetisierenden Eukaryoten wurde jedoch bislang nicht nachgewiesen. In der vorliegenden Arbeit wurde die Transkription mitochondrialer Gene durch die mitochondrialen Phagentyp-RNA-Polymerasen RpoTm und RpoTmp der Pflanze *Arabidopsis thaliana* untersucht.

Im mitochondrialen Genom von Arabidopsis wurden Transkriptionsstartpunkte ausgewählter Gene und Gencluster bestimmt. Erstmals wurde hier für eine dikotyle Pflanze gezeigt, dass mitochondriale Gene häufig multiple Promotoren besitzen. Die identifizierten Promotoren zeigten verschiedene Sequenzelemente und wichen zumeist signifikant von der für Dikotyle publizierten Konsensussequenz ab. Es wurde darüber hinaus die Funktion von Promotoren nachgewiesen, welche die Transkription nichtkodierender und vermutlich nicht funktioneller Sequenzen aktivieren. Zwischen Promotoraktivitäten in Blüten- und in Blattgewebe wurden keine qualitativen Unterschiede beobachtet. Architektur, Häufigkeit, Lokalisation und Nutzung mitochondrialer Promotoren implizieren eine wenig stringente Kontrolle der Transkriptionsinitiation in Arabidopsis-Mitochondrien.

Die Identifizierung mitochondrialer Promotoren in Arabidopsis ermöglichte die Rekonstitution eines definierten *in vitro*-Transkriptionssystems zur Analyse der Transkription mitochondrialer Gene durch RpoTm und RpoTmp. In *in vitr*o-Assays mit rekombinanten RNA-Polymerasen initiierte RpoTm an verschiedenen, jedoch nicht allen angebotenen Promotorsequenzen die Transkription, während für RpoTmp keine signifikante Promotorspezifität beobachtet wurde. Offenbar wird die Spezifität des mitochondrialen Transkriptionsapparates für zahlreiche Promotoren durch RpoTm vermittelt. RpoTm initiierte die *in vitro*-Transkription an Promotoren auf *supercoiled*-strukturierten, jedoch nicht auf linearen DNA-Molekülen und unterscheidet sich hierin von in früheren Arbeiten charakterisierten transkriptionsaktiven Extrakten aus Pflanzen-mitochondrien. Dieser Befund impliziert, dass Kofaktoren, deren Funktion bei einer *supercoiled*-Konformation der DNA nicht essentiell ist, in *in vitro*-Assays mit mitochondrialen Extrakten und *in vivo* die Aufschmelzung der DNA unterstützen. Im Genom von Arabidopsis wurde nach Genen potentieller Kofaktoren von Phagentyp-RNA-Polymerasen gesucht. Ein durch den Kernlocus At5g66360 (*MetA*) kodiertes Protein, dessen mitochondriale Lokalisation hier nachgewiesen wurde, zeigt Ähnlichkeit zu mtTFB, einem essentiellem Kofaktor der T7-ähnlichen RNA-Polymerasen in Hefe- und Säugermitochondrien. In *in vitr*o-Assays wurde die Initiation der RpoTm- und RpoTmp-abhängigen Transkription in keiner Weise durch MetA beeinflusst. Eine mtTFB-ähnliche Funktion von MetA in der mitochondrialen RNA-Synthese ist daher unwahrscheinlich. In Übereinstimmung hiermit wurde für MetA eine engere phylogenetische Beziehung zu nicht-mitochondrialen rRNA-Dimethylasen als zu mtTFB-Proteinen ermittelt.

Die hier vorgestellten Studien belegen die Transkription mitochondrialer Gene in Arabidopsis durch das Enzym RpoTm; für RpoTmp ist eine nicht-redundante mitochondriale Transkriptions-funktion denkbar, welche die Erkennung bekannter Promotoren nicht erfordert. Im Pflanzenreich wird hiermit erstmals der funktionelle Nachweis erbracht, dass ein nukleäres Phagentyp-RNA-Polymerasegen für ein Transkriptionsenzym mit mitochondrialer Promotorspezifität kodiert. Die Kofaktor-unabhängige *in vitro*-Spezifität von RpoTm für verschiedene Promotorsequenzen und die offenbar nicht stringente Initiationskontrolle *in vivo* legen nahe, dass eine individuelle Regulation mitochondrialer Gene in Arabidopsis auf Transkriptionsebene nicht erfolgt.

## SUMMARY

Mitochondria depend on a nucleus-encoded transcription machinery to express their genome and maintain mitochondrial function. In plants, nuclear T7 bacteriophage-like *RpoT* genes identified in a variety of species have been suggested to encode catalytic subunits of the mitochondrial transcription apparatus. Still, functional evidence has been lacking for RpoT enzymes being involved in mitochondrial transcription in photosynthetic eukaryotes. The present study examined the transcription of mitochondrial genes by two mitochondrial phage-type RNA polymerases encoded by the *RpoTm* and *RpoTmp* genes in the plant *Arabidopsis thaliana*.

To study *cis*-elements that are recognized by these enzymes, transcription initiation sites of selected mitochondrial genes and gene clusters in Arabidopsis were determined. Most genes were found to possess multiple promoters, revealing for the first time that promoter multiplicity is a common feature of mitochondrial genes in a dicotyledonous plant. The identified promoters displayed diverse sequence elements and for the most part deviated significantly from a nonanucleotide consensus derived previously for dicot mitochondrial promoters. Several promoters were moreover detected that activate transcription of non-coding and presumably non-functional sequences. No qualitative differences in promoter utilization were observed between leaves and flowers. Promoter architecture, distribution and utilization suggest a non-stringent control of transcription initiation in Arabidopsis mitochondria.

The knowledge of mitochondrial promoters in Arabidopsis allowed a defined *in vitro* transcription system to be set up in order to elucidate the roles of RpoTm and RpoTmp in mitochondrial transcription. Since RpoT-driven transcription from mitochondrial promoters possibly requires a complementation of RpoT enzymes with as yet unidentified auxiliary factors, the Arabidopsis genome was screened for potential cofactors of phage-type RNA polymerases. A protein encoded by the nuclear locus At5g66360 (*MetA*) and shown here to be imported into mitochondria displays sequence similarity to mtTFB, an essential cofactor of the mitochondrial T7-like RNA polymerase in yeast and mammals that is related to rRNA methyltransferases. *In vitro* transcription studies examined the abilities of recombinant RpoTm and RpoTmp to transcribe DNA from mitochondrial promoters, and moreover tested MetA for its potential to modulate the transcriptional performances of RpoTm and RpoTmp. RpoTm recognized a variety of promoters whereas RpoTmp displayed no significant promoter specificity. Sequence specificity of the mitochondrial transcription apparatus thus appears to be conferred by the RpoTm core enzyme for the majority of promoters. RpoTm differed in its transcriptional performance from formerly characterized plant mitochondrial extracts in that it did not specifically initiate transcription at promoters located on linear DNA but required supercoiled templates. This indicates a participation of (an) auxiliary factor(s) in DNA melting in transcription experiments using mitochondrial extracts and *in vivo*, which is obviated by a supercoiled DNA conformation. Transcription initiation by RpoTm or RpoTmp appeared to be not affected by the presence of MetA in the *in vitro* assay, indicating that MetA does not have a role equivalent to that of yeast or mammalian mtTFB in mitochondrial RNA synthesis. In line with this, phylogenetic analyses revealed MetA to be more closely related to a group of non-mitochondrial rRNA dimethylases than to fungal or animal mtTFBs.

The data presented here establish RpoTm to transcribe mitochondrial genes. They thus provide the first direct linkage in the plant kingdom between an RNA polymerase activity recognizing mitochondrial promoters and a nuclear gene encoding a mitochondrial phage-type RNA polymerase. Experimental results suggest RpoTmp to have a non-overlapping transcriptional role in mitochondria, which might not involve the recognition of known mitochondrial promoters. The cofactor-independent *in vitro* specificity of RpoTm for diverse promoter sequences and the apparently non-stringent control of transcription initiation *in vivo* imply that mitochondrial genes in Arabidopsis may not be regulated individually at the transcriptional level.

# I INTRODUCTION

## I.1 The mitochondrion, an endosymbiont-derived cell organelle

Mitochondria are the compartments of cellular respiration in eukaryotes and accommodate an energy-transducing system that generates ATP by coupling electron transport with oxidative phosphorylation (Saraste, 1999). The respiratory organelle is the descendant of a bacterial endosymbiont (Margulis, 1970; Margulis, 1981) and possesses its own vestigial genome. Analyses of the mitochondrial DNA (mtDNA) trace the evolutionary predecessors of mitochondria to a single ancestor whose closest contemporary relatives are found within the α division of the proteobacteria (Yang, et al., 1985). While the bacterial endosymbiont has long been considered to have been established in a nucleus-containing host cell, studies of unicellular eukaryotes have raised the possibility that the mitochondrion originated at essentially the same time as the nuclear compartment rather than in a subsequent event (Gray, et al., 1999). The majority of the original set of mitochondrial genes was either relocated to the nuclear genome or lost from the cell early in eukaryotic evolution (reviewed in Gray, 1992; Gray, et al., 1999). While animal mtDNA reduction appears to effectively have ceased in the common ancestor of all animals (Boore, 1999), gene transfer from the mitochondrion to the nucleus is an ongoing process in plants (Adams, et al., 2000; Adams, et al., 2002; Adams, et al., 1999). As a consequence of the unidirectional functional gene transfer, components participating in the diverse mitochondrial metabolic pathways and genetic processes largely are encoded in the nucleus and, following synthesis in the cytosol, are imported into the organelle (Herrmann, 2003; Martin and Herrmann, 1998). The protein complexes of the electron transport chain and mitochondrial ribosomes are assembled as mosaics of nucleus- and mitochondrion-encoded components (Figure 1). Correct assembly of these complexes therefore requires the coordinated expression of the mitochondrial and nuclear genomes. In photosynthetic eukaryotes, where the plastid as a second endosymbiont-derived organelle with a residual genome contributes to cellular processes, mitochondria need to coordinate gene function with yet another genetic compartment. The present study investigates as one essential element in organelle gene expression the transcription machinery in plant mitochondria.

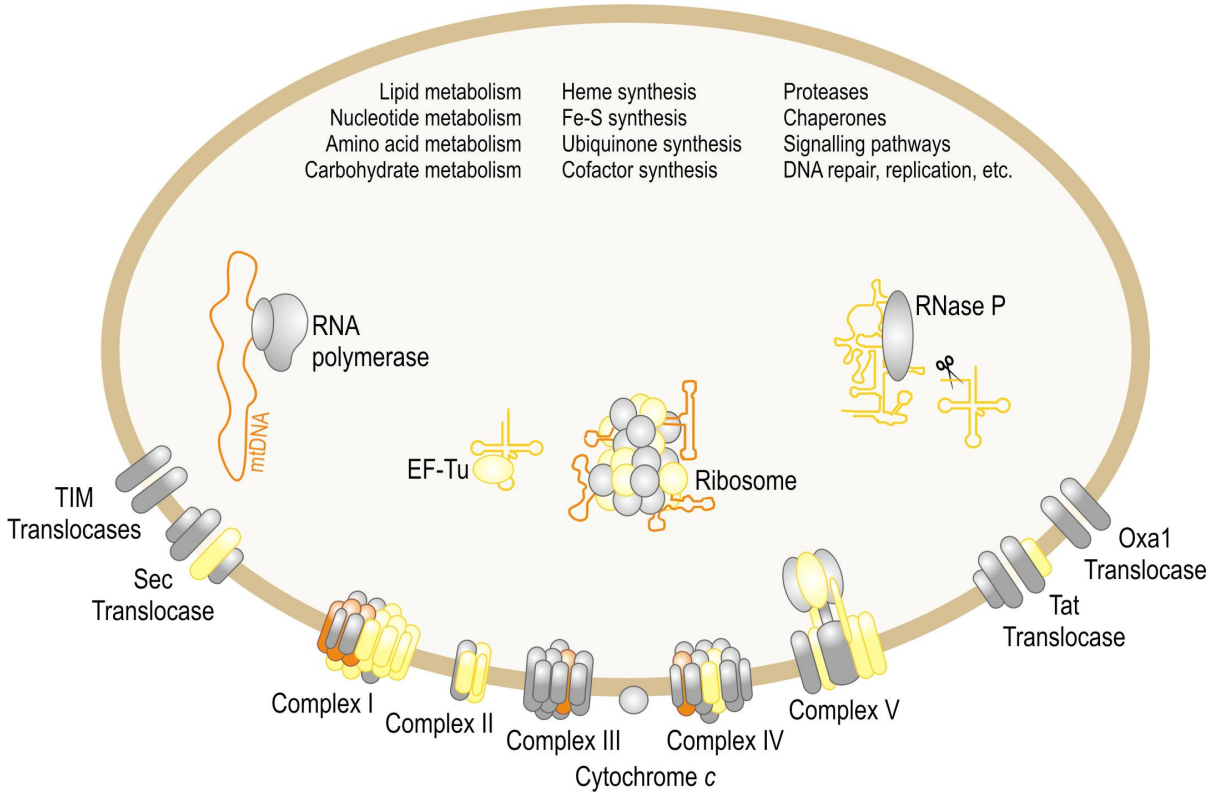



**Figure 1: Participation of mtDNA-encoded proteins and RNAs in biological processes in mitochondria.** The majority of mitochondrial functions are indicated; only the mitochondrial matrix and inner membrane are shown. Most mitochondrial components are nucleus-encoded (examples shown in grey), and the majority of mitochondrial processes have exclusively nucleus-encoded constituents (listed). Yellow symbols correspond to proteins and RNAs encoded by the mtDNA in some eukaryotes but by the nuclear genome in others. Very few components are specified by the mtDNA in all organisms (orange). Displayed components are involved in electron transport and oxidative phosphorylation (complexes I-V and cytochrome *c*), protein import and insertion into the inner membrane (TIM), protein export from the matrix and insertion into the inner membrane (Tat, Sec, Oxa1), mtDNA transcription (RNA polymerase), tRNA 5'-end processing (RNase P), protein synthesis (ribosomes and elongation factor EF-Tu). After (Burger, et al., 2003).

## I.2 Plant mitochondrial genomes

Plant mitochondrial genomes considerably vary in size but contain a fairly stable number of 50 to 60 genes (Handa, 2003; Kubo, et al., 2000; Notsu, et al., 2002; Sugiyama, et al., 2005; Unseld, et al., 1997). These may be dispersed or organized in gene clusters, and predominantly code for components of the respiratory chain and of the translational apparatus while the machinery that transcribes these genes is encoded in the nucleus. Mitochondrial genome size in land plants ranges from 187 kbp in *Marchantia polymorpha* to 570 kbp in *Zea mays* and may even reach up to 2400 kbp in certain *Cucurbitaceae* (Ward, et al., 1981). In contrast, animal mitochondria contain considerably smaller genomes, yet they do not encode a proportionately smaller number of genes (reviewed in Bullerwell and Gray, 2004; Burger, et al., 2003). For example, the mtDNA of *Arabidopsis thaliana* (~367 kbp; Figure 2) is 20 times larger than the human mitochondrial genome (~17 kbp) but codes for only approximately

twice the number of proteins and one more RNA than the human mtDNA (Anderson, et al., 1981; Unseld, et al., 1997). 33 proteins, three rRNAs and 20 tRNAs are encoded by the Arabidopsis mtDNA (Duchene and Marechal-Drouard, 2001; Unseld, et al., 1997), whereas the functional plant mitochondrion has been estimated to contain 1000 or more proteins (Millar, et al., 2004).

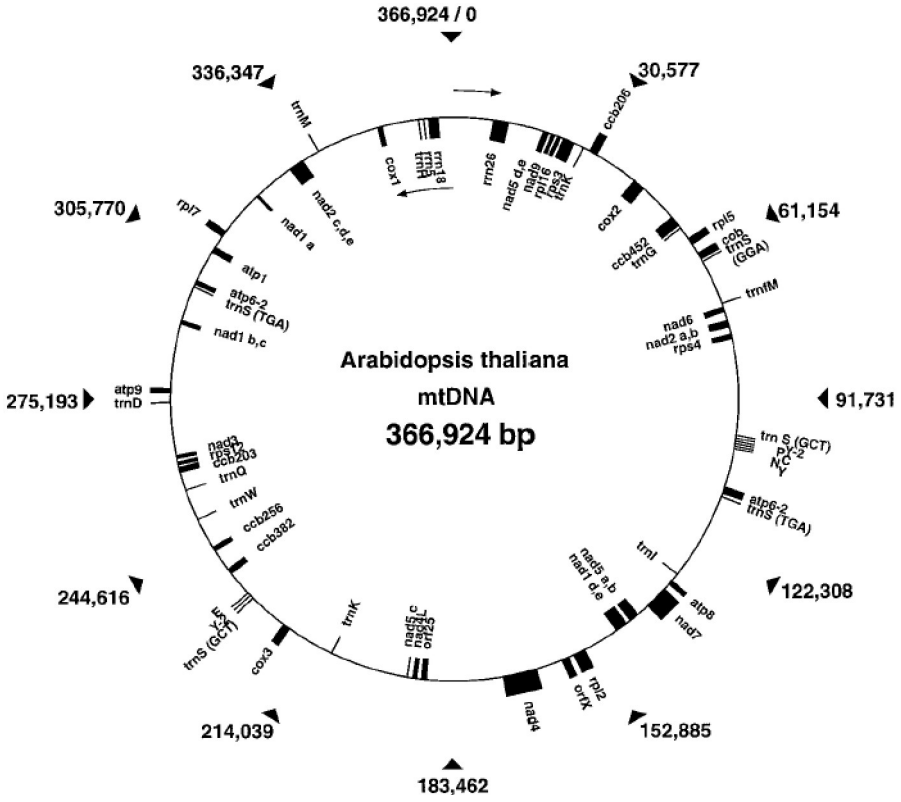



**Figure 2: Distribution of identified genes in the mitochondrial genome of Arabidopsis.** (Taken from Dombrowski, et al., 1998).

Mitochondrial genomes of angiosperms display large intergenic regions and have expanded through frequent duplications as well as by incorporating introns and DNA segments from the plastid and the nucleus in addition to sequences of as yet unknown origin (Marienfeld, et al., 1999; Palmer, et al., 2000). Rapid structural evolution of the plant mitochondrial genome, which is accounted for by recombination between the small and large repeats of the mtDNA, is contrasted by slow evolution in sequence (Palmer, 1990). Recombinational activity of repeated sequences moreover subdivides the mitochondrial genome into several different subgenomic DNA molecules (Andre, et al., 1992; Fauron, et al., 1995; Sugiyama, et al., 2005), the maintenance of which has in particular cases been shown to be under nuclear control (Abdelnoor, et al., 2003; Martinez-Zapater, et al., 1992). Low-frequency intragenic recombination events that are mediated by smaller repeats can result in

chimeric ORFs which, if transcriptionally active, may lead to pollen sterility (reviewed in Hanson and Bentolila, 2004; Linke and Börner, 2005).

Although plant mtDNA sequencing projects have commonly assembled a master circle from the complete genetic information (see Figure 2), the mitochondrial genome is maintained *in vivo* as not only subgenomic but also largely linear and branched molecules (Backert and Börner, 2000; Bendich, 1993; Oldenburg and Bendich, 1996; Oldenburg and Bendich, 2001; reviewed in Backert, et al., 1997).

DNA-containing fractions prepared from mung bean mitochondria have been characterized as membrane-associated chromatin-like nucleoids (Dai, et al., 2005). Isolated plant mitochondrial nucleoids were found to retain transcriptional activity and the ability to synthesize DNA *in vitro*, indicating that like in fungi and in mammals, mitochondrial nucleoids in plants are centres of mtDNA maintenance and expression (Dai, et al., 2005; Fey, et al., 1999). While proteins associated with plant mitochondrial nucleoids are still awaiting classification, a high mobility group (HMG) protein designated mtTFA has been identified as a highly abundant protein component of mtDNA-protein complexes in *Saccharomyces cerevisiae*, *Xenopus laevis* and humans (Alam, et al., 2003; Antoshechkin and Bogenhagen, 1995; Diffley and Stillman, 1992; Shen and Bogenhagen, 2001). Acting as a nucleoid architectural factor, mtTFA is required for mtDNA maintenance (Diffley and Stillman, 1991; Kanki, et al., 2004). Besides, mtTFA is an obligatory transcription factor in human mitochondria (see I.3.3.2). A candidate mitochondrial mtTFA-like protein in Arabidopsis has been suggested to be encoded by a sequence positioned in a cluster of nuclear genes that code for mitochondrial proteins involved in DNA and RNA metabolism (Elo, et al., 2003; Heinhorst, et al., 2004). Computational predictions of subcellular protein targeting are however not in support of this protein being a mitochondrial component (Elo, et al., 2003).

More progress has been made in characterizing nucleoid proteins in the plastid than in plant mitochondria (reviewed in Heinhorst, et al., 2004; Phinney and Thelen, 2005). While the prokaryotic histone-like DNA-binding protein HU is abundant in plastidial nucleoids of red algae, the major DNA-compacting protein in the plastids of plants has been identified as a 70-kDa sulfite reductase (SiR) (Sato, et al., 2001; Sato, et al., 2003). Addition of recombinant maize SiR has been demonstrated to increase the compaction of isolated plastid nucleoids *in vitro* and to concurrently repress *in vitro* transcription activity of nucleoids (Sekine, et al., 2002). DNA compaction through SiR is reversible and has been suggested to regulate the transcriptional activity in the chloroplast through changes in nucleoid compaction (Sekine, et

al., 2002). A protein similar in size to SiR was found to compact chloroplast nucleoids and suppress replication in soybean (Cannon, et al., 1999). It thus seems that the major nucleoid proteins of plant plastids and of yeast and animal mitochondria differ fundamentally in their effect on nucleoid activity.

## I.3 Transcription in higher plant mitochondria

### I.3.1 Mitochondrial promoters

*Cis*-regulatory elements of mitochondrial transcription in metazoa are confined to a discrete mtDNA region known as D-loop, and all genes are transcribed from one or two uni- or bidirectional promoters located in this region (reviewed in Shadel and Clayton, 1993; Tracy and Stern, 1995). In contrast, multiple promoters are active in yeast mitochondria, which share a sequence motif of nine nucleotides that is sufficient to efficiently promote transcription initiation *in vitro* (Tracy and Stern, 1995). Plant mitochondria similarly transcribe their genomes from numerous promoters (Tracy and Stern, 1995). Mitochondrial promoters have been analyzed in several plant species through identifying primary 5' termini of mitochondrial transcripts and aligning sequences surrounding transcription initiation sites, thereby revealing conserved promoter motifs (Figure 3; Fey and Marechal-Drouard, 1999; Hess and Börner, 1999). Moreover, sequence elements that are relevant for promoter function have been defined in *in vitro* transcription studies using complex mitochondrial extracts as a source of transcription activity (Binder, et al., 1995; Caoile and Stern, 1997; Dombrowski, et al., 1999; Hoffmann and Binder, 2002; Rapp, et al., 1993; Rapp and Stern, 1992). Sequences of up to 25 nucleotides around the transcription start site, which display the conserved motif YRTA (Y = T or C and R = A or G) immediately upstream of the initiating nucleotide, were found to be required for correct and efficient initiation of transcription *in vitro* (Hess and Börner, 1999). The majority of higher plant mitochondrial promoters exhibit an A/T-rich sequence element immediately upstream of the promoter core, which has been proven essential for the full function of different dicot and monocot mitochondrial promoters *in vitro* (Dombrowski, et al., 1999; Rapp, et al., 1993).

In mitochondrial promoters of dicotyledonous plants, the YRTA core motif is embedded in an extended consensus of nine nucleotides, CRTAAGAGA, with the initiating nucleotide at the penultimate position (Figure 3; Binder, et al., 1996). Only a few transcription start sites in dicot mitochondria coincide with sequences lacking a recognizable core motif (Binder, et al., 1994; Brown, et al., 1991). In contrast to mitochondrial promoters of *Oenothera berteriana* and potato that conform to the nonanucleotide consensus and are recognized by a pea *in vitro*

transcription system, deviating mitochondrial promoters of both species do not function in the heterologous system, which implies that specific mechanisms are involved in transcription initiation at alternative promoters (Binder, et al., 1995). Moreover, studies on mitochondrial transcription start sites in both dicots and monocots support the idea that the utilization of particular promoters requires a distinct nuclear background (Edqvist and Bergman, 2002; Newton, et al., 1995).


**Figure 3: Conserved sequence elements of plant mitochondrial promoters.** Consensus promoter sequences in monocots and dicots are shown essentially as deduced by (Fey and Marechal-Drouard, 1999). The YRTA core motif is written bold, and initiating nucleotides are underlined. Conserved elements of the central promoter domain are highlighted orange; A/T-rich upstream elements are indicated. Upper case nucleotides denote highly conserved positions (frequency of appearance >75%); lower case letters indicate weakly conserved nucleotides (frequency between 50 and 75%). R = A or G; W = A or T; K = T or G; Y = C or T; n = any nucleotide.

Mitochondrial promoters of monocotyledonous plants often deviate in the YRTA core motif and are overall less conserved than dicot mitochondrial promoters (Fey and Marechal-Drouard, 1999; Hess and Börner, 1999). Most promoters comprise a central domain around the transcription initiation site, which contains the YRTA tetranucleotide at varying distance from the initiating nucleotide (Figure 3; Caoile and Stern, 1997; Covello and Gray, 1991; Rapp, et al., 1993). Among the few promoters displaying no conserved motif is the alternative promoter *cpc* preceding the *cox2* coding sequence in *Zea perennis* (Newton, et al., 1995). Mitochondrial genes have been described to display multiple transcription initiation sites in monocots far more frequently than in dicots (Lupold, et al., 1999; Mulligan, et al., 1988). It has been suggested that promoter multiplicity is maintained in order to ensure mitochondrial gene expression despite frequent genome rearrangements (compare I.2; Lupold, et al., 1999); alternatively, multiple transcription initiation sites may merely be a consequence of a promiscuous mitochondrial transcription machinery (Lupold, et al., 1999).

### I.3.2 Mitochondrial T7 bacteriophage-like RNA polymerases

#### I.3.2.1 Plant RpoT genes encoding phage-type transcriptases

Evolution of the mitochondrion was in nearly all organisms accompanied by the loss of genes encoding the bacterial-type RNA polymerase and the acquisition of a different transcription apparatus, the protein components of which are encoded in the nucleus and imported into the organelle (reviewed in Gray and Lang, 1998; Hess and Börner, 1999; Tracy and Stern, 1995). Excepting the brown alga *Pylaiella littoralis*, none of the mitochondrial genomes of photosynthetic eukaryotes sequenced to date harbours sequence motifs of bacterial-type $\sigma^{70}$-dependent promoters. Instead, promoters of diverse architecture have been identified (see I.3.1). Mitochondrial RNA polymerases accordingly differ from enzymes of the bacterial type.

In *Saccharomyces cerevisiae*, the nuclear *RPO41* gene encodes a phage-type RNA polymerase operating as catalytic subunit of the mitochondrial transcription machinery (Greenleaf, et al., 1986; Masters, et al., 1987; see Figure 4). A phage-type enzyme was moreover shown to function as core RNA polymerase in mitochondrial transcription in humans (Falkenberg, et al., 2002; Tiranti, et al., 1997). DNA sequences homologous to mitochondrial phage-type and bacteriophage T3/T7 RNA polymerases have been amplified from a phylogenetically broad range of multicellular and unicellular eukaryotes, suggesting that a phage-type enzyme was recruited to function in mitochondrial transcription at an early stage in the evolution of the mitochondrion (Cermakian, et al., 1996).

Genes encoding T3/T7 phage-like RNA polymerases, which are commonly designated *RpoT* genes, have been identified in the nuclear genomes of various angiosperms such as *Chenopodium album* (Weihe, et al., 1997), Arabidopsis (Hedtke, et al., 1997; Hedtke, et al., 2000), maize (Chang, et al., 1999; Young, et al., 1998), wheat (Ikeda and Gray, 1999), *Nicotiana tabacum* (Hedtke, et al., 2002), *Nicotiana sylvestris* (Kobayashi, et al., 2002; Kobayashi, et al., 2001), barley (Emanuel, et al., 2004), and in the moss *Physcomitrella patens* (Kabeya, et al., 2002; Richter, et al., 2002). Homologous sequences were moreover detected in the gymnosperm *Pinus taeda* (U. Richter, HU Berlin, personal communication) and in green algae (A. Weihe, HU Berlin, personal communication).

According to *in vitro* and *in vivo* import studies, a small family of three *RpoT* genes in Arabidopsis encodes a mitochondrial RNA polymerase (RpoTm), a plastidial enzyme (RpoTp) as well as a polypeptide imported into both mitochondria and plastids (RpoTmp; see Figure 4) (Hedtke, et al., 1997; Hedtke, et al., 2000; Hedtke, et al., 1999). Comparable

subcellular targeting of *RpoT* gene products is observed in *Nicotiana* species (Hedtke, et al., 2002; Kobayashi, et al., 2001). In contrast, monocots have so far been determined to harbour no more than two *RpoT* genes, of which one codes for a mitochondrial RNA polymerase and the other for a plastidial enzyme (Figure 4; Chang, et al., 1999; Ikeda and Gray, 1999). Dual targeting of *RpoT* gene products to both mitochondria and plastids being the result of two different translational starts has been reported for the two phage-type RNA polymerases genes *RpoT1* and *RpoT2* of *Physcomitrella* (Figure 4; Richter, et al., 2002). For both *RpoT1* and *RpoT2*, translation initiation at the first of two in-frame AUG start codons was found to yield a polypeptide that is targeted to plastids, whereas initiation at the downstream AUG gave rise to a mitochondrial protein (Richter, et al., 2002). *In vivo* translation initiation at the first AUG and a plastidial localization of both RpoT1 and RpoT2 in *Physcomitrella* and of RpoTmp in Arabidopsis have recently been questioned by (Kabeya and Sato, 2005) and await further experimental proof.

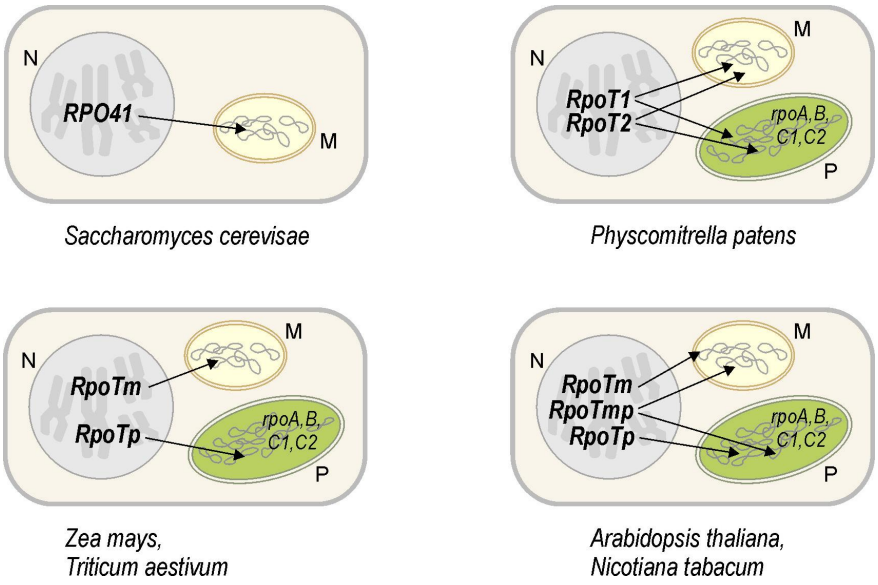



**Figure 4: Nuclear genes encoding organellar phage-type RNA polymerases.** Genes in the nucleus (N, grey) of eukaryotic organisms code for T7 phage-like RNA polymerases which, following their synthesis in the cytoplasm, are imported into mitochondria (M, yellow) and plastids (P, green).

The small *RpoT* gene families found in *Physcomitrella* and in higher plants appear to be the result of independent gene duplication events dating after the separation of bryophytes from the vascular plant lineage (Richter, et al., 2002). Phylogenetic analyses have shown all higher plant RpoTp enzymes to form a sister clade to the group of RpoTm and RpoTmp enzymes (Richter, et al., 2002). Interestingly, preliminary phylogenetic analyses are indicative

of a closer relationship of monocot RpoTm to dicot RpoTmp than to dicot RpoTm (U. Richter, HU Berlin, personal communication).

A DNA polymerase or reverse transcriptase has been proposed as predecessor of a single-subunit RNA polymerase that was the common ancestor of phage-encoded and nucleus-encoded phage-type RNA polymerases (Cermakian, et al., 1997). The ancestral single-subunit RNA polymerase gene has moreover been suggested to have originated at essentially the same time as the mitochondriate eukaryotic cell (Cermakian, et al., 1997). The discovery of cryptic prophages related to T3/T7 bacteriophages in several genomes of proteobacteria has inspired a modified scenario: A prophage that encoded among other phage proteins a T3/T7-like RNA polymerase existed in the bacterial endosymbiont being the predecessor of the mitochondrion (Filee and Forterre, 2005). Transfer of the prophage genes to the nucleus may have resulted in the reactivation of these formerly silent genes (Filee and Forterre, 2005).

### I.3.2.2 Roles of RpoT enzymes

Distinct functions of RpoTm and RpoTmp in mitochondria and of RpoTp and RpoTmp in plastids of dicots are yet to be assigned. The *RpoTm* and *RpoTmp* genes in Arabidopsis have been reported to display overlapping expression patterns in different tissues and at different developmental stages (Emanuel, et al., 2005). Emanuel et al. (2005) therefore suggested RpoTm and RpoTmp to recognize different types of mitochondrial promoters. A contrasting picture of RpoTm and RpoTmp functions has been stimulated by studies of a transgenic Arabidopsis line carrying a T-DNA insertion in the *RpoTmp* gene (Baba, et al., 2004). Transgenic plants displayed no apparent alterations compared to wild-type individuals in mitochondrial transcript accumulation. Based predominantly on the observation that in the mutant, the induction of several plastid genes in dark-grown seedlings upon illumination was delayed, Baba et al. (2004) proposed RpoTmp to be the key RNA polymerase transcribing organellar genes during early seedling development and favoured a role of both RpoTm and RpoTp at a later developmental stage.

Only indirect evidence has been provided that the mitochondrial phage-type RNA polymerases encoded by *RpoT* genes in higher plants have a role in transcription of mitochondrial genes. No sequences that might encode potential mitochondrial RNA polymerases of known enzyme structure have been traced in the fully sequenced Arabidopsis genome besides the previously characterized genes *RpoTm* and *RpoTmp* (Hedtke, et al., 1997; Hedtke, et al., 2000; The Arabidopsis Genome Initiative, 2000). The conservation of functionally critical amino acid positions of the T7 enzyme (McAllister and Raskin, 1993;

Sousa, et al., 1993) in RpoTm and RpoTmp as well as in other plant RpoT enzymes argues for their transcriptional function (Hess and Börner, 1999; see I.3.2.3 and Figure 5). Moreover, recombinant RpoTm and RpoTmp were shown to non-specifically transcribe DNA *in vitro* (Hedtke, et al., 2000; Kühn, 2001).

The mitochondrial RNA polymerase functions as primase for mtDNA replication in humans and presumably also in yeast (reviewed in Shadel and Clayton, 1997; Tracy and Stern, 1995). A similar role of phage-type RNA polymerases and association of origins of replication with mitochondrial transcription start sites in plants remain to be demonstrated.

### I.3.2.3 Structure of bacteriophage and phage-type RNA polymerases

The RNA polymerase of the T7 bacteriophage is a 99-kDa single-polypeptide-chain enzyme that is able to recognize specific promoter sequences, correctly initiate transcription and catalyze transcript elongation (reviewed in Steitz, 2004). Although the phage-type RNA polymerases of eukaryotic organisms most probably require auxiliary proteins to initiate transcription at organellar promoters (see I.3.3), the thoroughly studied T7 enzyme commonly serves as a model for both bacteriophage and eukaryotic phage-type RNA polymerases.

Comparisons between amino acid sequences of phage- and nucleus-encoded enzymes have identified conserved domains comprising identical or conservatively substituted positions (Chang, et al., 1999; Hedtke, 1998). Catalytically relevant structures of the T7 RNA polymerase are formed through amino acids of the C-terminal half of the polypeptide (McAllister and Raskin, 1993; Sousa, et al., 1993). Crystal structures of the T7 RNA polymerase have revealed this portion of the protein to fold into the “fingers”, “palm” and “thumb” subdomains (Jeruzalmi and Steitz, 1998; Sousa, et al., 1993; see Figure 5) that are typical to members of a superfamily of nucleic acid polymerases also including certain DNA polymerases and reverse transcriptases (reviewed in Sousa, 1996). The similarity of plant organellar RNA polymerases to the T7 enzyme is most apparent for sequence regions corresponding to palm and fingers, which include all essential residues that participate in catalysing RNA synthesis (Figure 5; Hess and Börner, 1999), and references therein). In contrast, elements contributing to promoter recognition by phage RNA polymerases are poorly conserved in plant phage-type transcriptases, possibly reflecting the divergence in architecture between promoters recognized by the two groups of enzymes, as well as different compositions of initiating RNA polymerase complexes (Chang, et al., 1999; Ikeda and Gray, 1999; Jeruzalmi and Steitz, 1998; see I.3.3). Based on structural modelling, Yeast Rpo41 has been suggested to possess two regions that correspond in structure, yet not in sequence, to the

"specificity loop" and intercalating β-hairpin involved in promoter recognition and DNA melting by T7 RNA polymerase (Matsunaga and Jaehning, 2004). In the T7 RNA polymerase-promoter complex, an antiparallel β-hairpin referred to as specificity loop specifically interacts via hydrogen-bonding with bases of the promoter sequence (Cheetham, et al., 1999; Rong, et al., 1998). The intercalating β-hairpin, which is part of the N-terminal domain of the T7 enzyme, facilitates melting of the promoter duplex (Cheetham, et al., 1999). Rpo41 was recently shown to accurately initiate promoter-specific transcription *in vitro* from supercoiled and pre-melted DNA templates in the absence of the obligatory yeast mitochondrial transcription factor sc-mtTFB, indicating that promoter specificity determinants indeed reside in the Rpo41 polypeptide rather than in sc-mtTFB (Matsunaga and Jaehning, 2004, see I.3.3.1).

Differences in size between nucleus-encoded phage-type RNA polymerases are primarily accounted for by varying N-terminal extensions. Including transit peptides, T7-like transcriptases of plants are proteins of around 110 kDa, whereas the yeast and human mitochondrial RNA polymerases are 145 and 130 kDa in size (Hess and Börner, 1999), and references therein). The N-terminal extension of yeast Rpo41 and a C-terminal insertion found only in the yeast enzyme are required for stable mtDNA maintenance and represent independent functional domains that may have been acquired through gene fusion events (Lisowsky, et al., 2002; Wang and Shadel, 1999). In heterologous complementation experiments, Arabidopsis RpoTm and various Rpo41/RpoTm chimeras were unable to functionally substitute for Rpo41 *in vivo* (Lisowsky, et al., 2002).

A number of separate crystal structures have depicted T7 RNA polymerase at different stages from promoter binding to elongation (Cheetham, et al., 1999; Cheetham and Steitz, 1999; Tahirov, et al., 2002; Yin and Steitz, 2002; Yin and Steitz, 2004). Following promoter rcognition, duplex DNA opening and repeated abortive initiation attempts, a major conformational change in the N-terminal domain removes the promoter-binding site and creates a tunnel for the transcript to pass through during the elongation phase in which the enzyme completes the RNA product processively without dissociation until termination (Steitz, 2004). Transcription termination signals recognized by the T7 RNA polymerase have been characterized (He, et al., 1998; Lyakhov, et al., 1998; Macdonald, et al., 1994), whereas no such sequences have been described in organelles. Mitochondrial transcripts in plants are instead considered to be terminated through the action of nucleases that define RNA 3' ends (Dombrowski, et al., 1997; Perrin, et al., 2004).

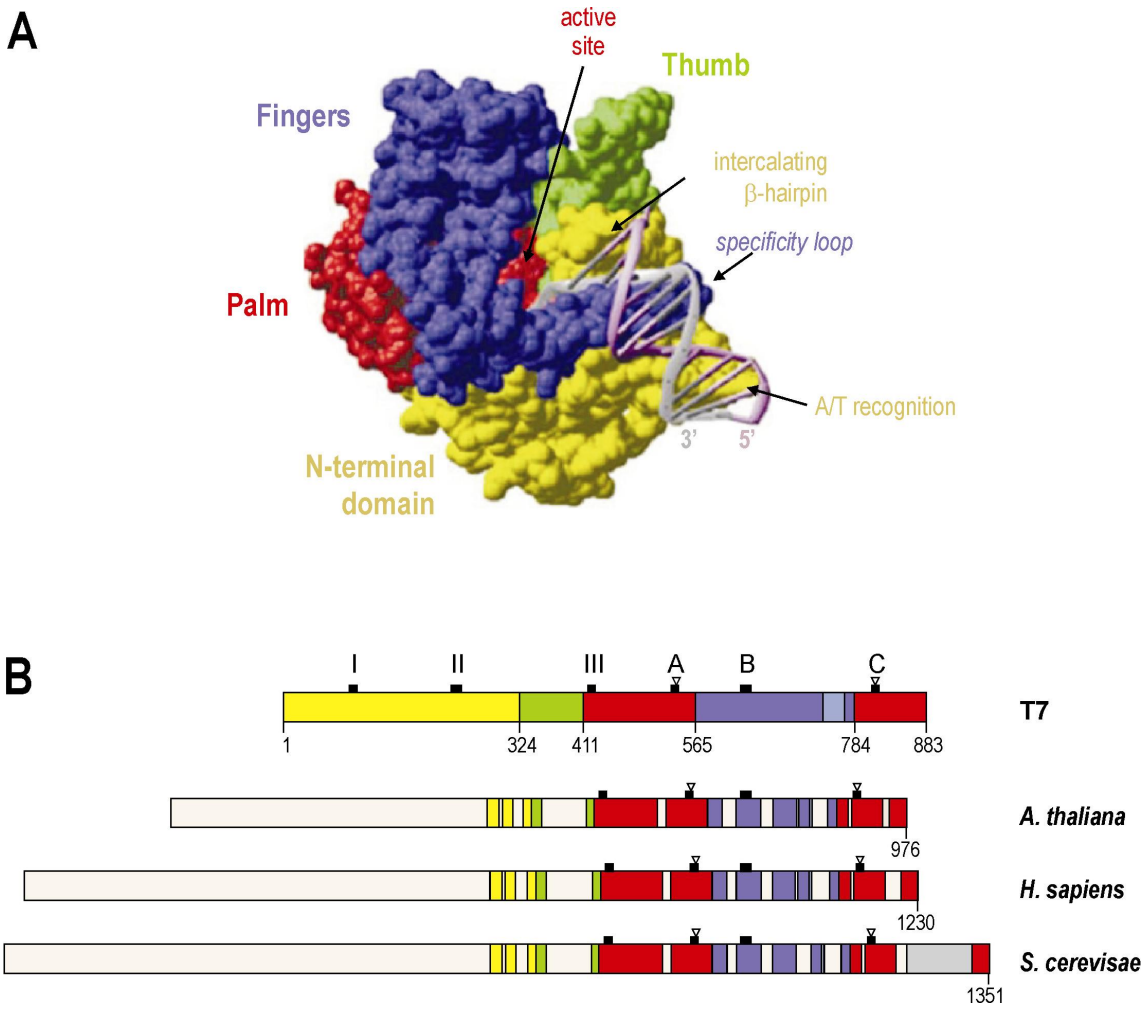



**Figure 5: Conserved sequence regions of the T7 phage and eukaryotic phage-type RNA polymerases.** (**A**) Surface representation of the T7 RNA polymerase-promoter complex structure specifying the N-terminal domain (yellow) and the RNA polymerase subdomains "thumb" (green), "palm" (red) and "fingers" (blue; from (Cheetham, et al., 1999). (**B**) Amino acid sequence organization of T7 RNA polymerase domains (colours as in Figure 5A) and homologous amino acid sequence regions of organellar phage-type RNA polymerases, based on a sequence comparison derived in (Hedtke, 1998). Arabidopsis RpoTm is shown as one member of the group of mitochondrial and plastidial enzymes of land plants which do not greatly differ in their sequence organization and size of the N-terminus. Black squares mark the positions of the motifs T/DxxGR (III), A, B and C (according to (Delarue, et al., 1990) that are important for RNA polymerase function and conserved in all enzymes, and of a T7-specific β-hairpin (I) and a structure involved in recognition of A/T-rich promoter regions (II). Open triangles denote the invariant residues Asp537 and Asp812 acting as ligands to two catalytic $Mg^{2+}$ ions at the RNA polymerase active site (Woody, et al., 1996). The region corresponding to the specificity loop of the fingers subdomain is depicted in light blue; a yeast-specific C-terminal insertion is shown in grey.

### I.3.3 Mitochondrial transcription factors

Unlike the single-subunit RNA polymerases of bacteriophages, mitochondrial phage-type RNA polymerases require auxiliary factors to initiate transcription at promoter sequences. To date, mitochondrial transcription factors have been characterized in *S. cerevisiae* (e.g. (Matsunaga and Jaehning, 2004; Schinkel, et al., 1987; Winkley, et al., 1985), *X. laevis* (Bogenhagen, 1996; Bogenhagen and Insdorf, 1988), *Drosophila melanogaster* (Matsushima, et al., 2005; Matsushima, et al., 2004), humans (Falkenberg, et al., 2002; Fisher and Clayton,

1988; McCulloch, et al., 2002) and mouse (Gaspari, et al., 2004). While no such proteins have yet been isolated from mitochondria of photosynthetic eukaryotes, a number of studies suggest transcriptional cofactors to function in mitochondrial transcription in plants. An essential role of promoter-specific cofactors in plant mitochondria gained support from analyses of mitochondrial transcription in maize (Young and Lonsdale, 1997). A nucleus-encoded factor has moreover been suggested to be involved in transcription in *Z. perennis* mitochondria where initiation at the *cox2* promoter *cpc* was found to depend on the presence of the dominant allele of the nuclear *MCT* gene (Newton, et al., 1995). Attempts to fractionate transcriptionally active mitochondrial extracts resulted in the loss of promoter specificity, indicating that several loosely associated proteins may contribute to promoter recognition (Binder and Brennicke, 2003). Searches for cofactors of mitochondrial phage-type RNA polymerases in higher plants are not unlikely to identify proteins with similarity to mitochondrial transcription factors in yeast and animals, which complement the same type of enzyme (compare I.3.2).

### I.3.3.1 Yeast and animal mtTFB

In addition to the core RNA polymerase Rpo41, transcription initiation at mitochondrial promoters in *S. cerevisiae* requires a single accessory protein of 43 kDa first described as Mtf1 (Figure 6; Lisowsky and Michaelis, 1988; Schinkel, et al., 1987) and also referred to as sc-mtTFB (Shadel and Clayton, 1993). While neither Rpo41 nor sc-mtTFB was able on its own to specifically interact with promoter sequences in gel mobility shift assays (Schinkel, et al., 1988), the non-specifically transcribing core was found to recognize mitochondrial promoters on a linear DNA template when complemented with sc-mtTFB in *in vitro* transcription experiments (Schinkel, et al., 1987). The latter was therefore considered the specificity factor of Rpo41 (Schinkel, et al., 1987). Functional studies of mutant variants of sc-mtTFB motivated the alternative view that specificity determinants of promoter recognition reside in the core enzyme, and that sc-mtTFB may play a role in unwinding DNA during transcription initiation (Shadel and Clayton, 1995). Intrinsic promoter specificity of the core RNA polymerase was later confirmed not only for yeast Rpo41 (Matsunaga and Jaehning, 2004) but in part also for the human and mouse core enzymes (Gaspari, et al., 2004). On supercoiled or premelted DNA templates, Rpo41 alone was able to accurately initiate transcription *in vitro* at mitochondrial promoters (Matsunaga and Jaehning, 2004). While sc-mtTFB was obligatory for specific transcription of linear templates, addition of sc-mtTFB stimulated specific transcription from supercoiled DNA but inhibited escape into

productive elongation from a pre-melted promoter (Matsunaga and Jaehning, 2004). The authors concluded that sc-mtTFB facilitated DNA melting, but not promoter recognition, possibly by inducing and stabilizing structural changes in Rpo41 that enable the enzyme to open the double helix and form an open promoter complex.

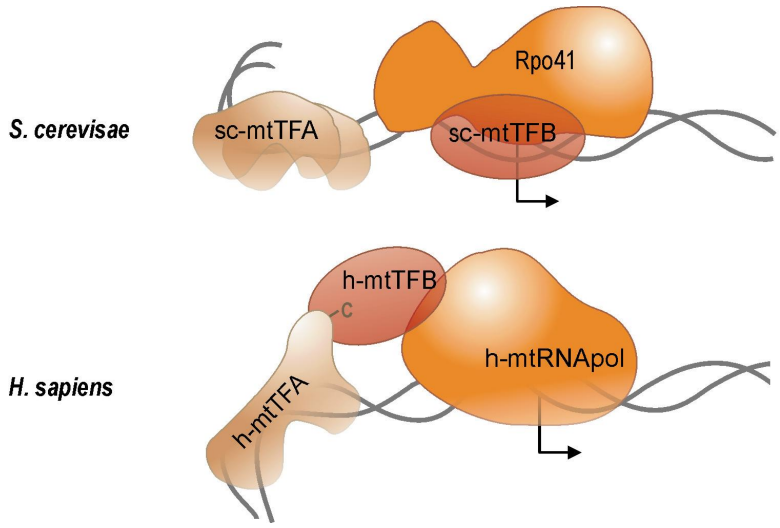



**Figure 6: Components of the transcription machineries in yeast and human mitochondria.** Yeast transcription initiation model after (Shadel and Clayton, 1993); transcription initiation in human mitochondria as proposed by McCulloch and Shadel (2003). See text for details on mtTFB and mtTFA functions and cofactor interactions with the mitochondrial phage-type RNA polymerases Rpo41 and h-mtRNApol. Cofactor designations adhere to the nomenclature suggested by Shadel and Clayton (1993). mtTFA-induced DNA bending is indicated; bent arrows mark transcriptional starts.

Analyses of the RNA polymerase composition before, during and shortly after transcription initiation indicated that the Rpo41 core and sc-mtTFB form a holoenzyme in solution prior to DNA binding and promoter recognition (Mangus, et al., 1994). Upon binding to the promoter, the RNA polymerase holoenzyme induces a bend in the DNA, which has been shown to enhance promoter activity *in vitro* (Schinkel, et al., 1988). Following transcription initiation, the factor dissociates from the catalytic subunit and is available for subsequent initiation events (Mangus, et al., 1994). The observation that sc-mtTFB behaved like bacterial sigma factors during transcription initiation stimulated a series of experiments investigating functional and structural similarities between sigma factors and sc-mtTFB (Cliften, et al., 2000; Cliften, et al., 1997; Jang and Jaehning, 1991). However, resolution of the three-dimensional structure of sc-mtTFB later revealed this mitochondrial transcription factor to structurally resemble *S*-adenosyl-L-methionine (SAM)-dependent rRNA dimethylases such as the *Bacillus subtilis* rRNA methyltransferase ErmC' (Schubot, et al., 2001).

Sequence similarity, albeit limited, of mitochondrial transcription factors to rRNA dimethylases became apparent following the identification of two sequences encoding mitochondrial sc-mtTFB-like factors in humans (Falkenberg, et al., 2002; McCulloch, et al., 2002). Homologues to human mtTFB1 and mtTFB2 (referred to as h-mtTFB1 and h-mtTFB2) where subsequently detected in mouse (Rantanen, et al., 2003) and *Drosophila* (Matsushima, et al., 2005; Matsushima, et al., 2004). Both h-mtTFB1 and h-mtTFB2 were found to form a stable heterodimer with human mitochondrial RNA polymerase and to induce transcription initiation at the human mtDNA promoters LSP and HSP *in vitro* in the presence of h-mtTFA (Figure 6; Falkenberg, et al., 2002), an additional essential cofactor of human mitochondrial transcription (see I.3.3.2).

Despite poor overall sequence similarity of sc-mtTFB to ErmC', there is extensive structural agreement between the two proteins around the SAM-binding site, and amino acid residues involved in SAM binding are conserved in sc-mtTFB (Schubot, et al., 2001). Therefore, Schubot et al. (2001) suggested that sc-mtTFB could bind SAM and methylate the nascent RNA chain *in vivo*. Alternatively, an RNA-modifying enzyme may have evolved to function solely as transcription factor. While no methyltransferase activity of sc-mtTFB or h-mtTFB2 has been demonstrated so far, h-mtTFB1 has been reported to bind SAM and to substitute for the *Escherichia coli* rRNA dimethylase KsgA in methylating the 16S rRNA at a conserved stem-loop (McCulloch, et al., 2002; Seidel-Rogol, et al., 2003). The homologous 12S rRNA in human mitochondria was moreover shown to be similarly modified at this site (Seidel-Rogol, et al., 2003). Analyses of transcriptionally competent h-mtTFB1 variants carrying point mutations in conserved methyltransferase motifs indicated that the ability of the protein to function as transcription factor *in vitro* is independent of SAM binding and methyltransferase activity (McCulloch and Shadel, 2003). The authors therefore concluded h-mtTFB1 to be a bifunctional protein with two separable activities.

### I.3.3.2 Yeast and animal mtTFA

In order to accurately and efficiently initiate transcription at the LSP and HSP promoters of the human mtDNA, the heterodimer composed of h-mtTFB and core RNA polymerase depends on the presence of the cofactor h-mtTFA (Falkenberg, et al., 2002; Fisher and Clayton, 1985; Fisher and Clayton, 1988). h-mtTFA is a protein of 25 kDa that comprises two HMG boxes in tandem (Parisi and Clayton, 1991). HMG boxes have been characterized as functional domains of DNA-binding proteins and interact with the minor grove of the double helix primarily at sites of unusual DNA conformation, where they induce dramatic bending

and structural deformation of the DNA (Antoshechkin, et al., 1997; Giese, et al., 1997; Grosschedl, et al., 1994; Love, et al., 1995). h-mtTFA specifically binds to sequences upstream of the transcription initiation sites, which have been characterized as distal promoter elements (Fisher, et al., 1987; Gaspari, et al., 2004). Specific promoter binding by h-mtTFA as well as exact spacing between distal promoter elements and transcription start sites were shown to be crucial for transcription activation (Dairaghi, et al., 1995).

A h-mtTFA homologue is abundantly found in mitochondria of *S. cerevisiae* (Diffley and Stillman, 1991). sc-mtTFA is dispensable for transcription initiation at promoters of the yeast mtDNA (Xu and Clayton, 1992) and instead appears to play a major role in structural organization and stable maintenance of the mtDNA (Diffley and Stillman, 1991). Like h-mtTFA, sc-mtTFA is able to specifically associate with regulatory sequences of the mtDNA and also non-specifically bind DNA, and mediates condensation and unwinding as well as bending of the double helix (Diffley and Stillman, 1992; Fisher, et al., 1992). Transcription initiation by the RNA polymerase holoenzyme is slightly stimulated by sc-mtTFA (Parisi, et al., 1993). Presumably, binding of sc-mtTFA to the mtDNA leads to a favourable exposition of *cis*-regulatory elements to the transcription apparatus (Diffley and Stillman, 1992).

The capacity of h-mtTFA to act as efficient and promoter sequence-specific transcriptional activator has been attributed to a C-terminal extension of the factor and a linker peptide between the two HMG boxes, which are both lacking in sc-mtTFA (Dairaghi, et al., 1995). Transfer of the C-terminal tail of h-mtTFA onto the yeast factor was shown to turn sc-mtTFA into a transcription factor activating the human mtDNA promoter LSP (Dairaghi, et al., 1995). Thus, h-mtTFA function appears to have evolved through the acquisition of novel structural domains. The C-terminal tail of h-mtTFA has been proposed to interact with h-mtTFB, thereby positioning the heterodimer composed of h-mtTFB and core RNA polymerase at the transcription initiation site demarcated by a specific h-mtTFA/promoter complex (Figure 6, McCulloch and Shadel, 2003). Corroborating this model, promoter selectivity of the mouse and human transcription machineries has been dissected into binding of mtTFA to distal promoter elements and specificity of the core enzyme for nucleotides proximal to the transcription initiation site (Gaspari, et al., 2004).

### I.3.3.3 Mitochondrial transcription factors in plants

To date, no mtTFA or mtTFB homologues have been isolated from plant mitochondria, and the function of such proteins in plant organelles is unclear. Attempts to detect HMG box proteins in maize mitochondrial extracts using a sc-mtTFA antibody did not identify mtTFA

candidates (Andrea T. Descheneau, University of Missouri, Columbia, USA; personal communication). Moreover, application of a protocol that had been successfully employed for the preparation of mtTFA from yeast and human mitochondria failed to purify homologous proteins from pea mitochondria (Däschner, et al., 2001; Hatzack, et al., 1998).

A biochemical approach directed at isolating transcription factors from wheat mitochondria lead to the identification of p63, a 63-kDa protein described to specifically bind to the yeast *cox2* promoter (Ikeda and Gray, 1999). Addition of recombinant p63 to a transcriptionally active extract prepared from wheat mitochondria, which per se accurately initiated transcription at the *cox2* promoter, appeared to stimulate transcription from this promoter *in vitro* (Ikeda and Gray, 1999). The authors therefore suggested p63 to play a role in mitochondrial transcription and moreover pointed out a limited amino acid sequence similarity of p63 to sc-mtTFB. However, p63 later emerged to be a member of the large family of organellar PPR proteins and may rather be involved in posttranscriptional processes (Lurin, et al., 2004; Small and Peeters, 2000; J. Gualberto, IBMP CNRS, Strasbourg, France, personal communication).

#### I.3.3.4 Cofactors of phage-type RNA polymerases in plastids

The homology of nucleus-encoded plastid and mitochondrial RNA polymerases in plants and the presence of yet another phage-type RNA polymerase in both mitochondria and plastids in dicotyledonous plants (Chang, et al., 1999; Hedtke, et al., 1997; Hedtke, et al., 2000; Hedtke, et al., 2002; Ikeda and Gray, 1999), as well as the similarity of promoters that are recognized by these enzymes (Weihe and Börner, 1999) raise the question whether the two organelles harbour similar transcriptional cofactors interacting with these core RNA polymerases. The characterization of auxiliary factors in the plastid may well aid the identification of such proteins in the mitochondrion. CDF2, a DNA-binding factor isolated from spinach chloroplast, has been reported to stimulate transcription of the *rrn* operon by a nucleus-encoded RNA polymerase activity (Bligny, et al., 2000). However, structural details of CDF2 have hitherto escaped revelation (Bligny, et al., 2000), and no CDF2-like activity has been purified so far from plant mitochondria. Out of six nucleus-encoded sigma factors imported into Arabidopsis plastids, the protein designated Sig2 was found to additionally localize to mitochondria in GFP import assays (Tandara, 2000). Moreover, the maize orthologue Sig2B was determined by immunoblot analyses to co-purify with both mitochondria and plastids (Beardslee, et al., 2002). Yet, available experimental data so far relate Sig2 function to the bacterial-type plastidial RNA polymerase rather than phage-type

enzymes in mitochondria or plastids (Beardslee, et al., 2002; Kanamaru and Tanaka, 2004, and references therein).

A regulatory role has recently been deduced of the plastid-encoded $tRNA^{Glu}$ in plastidial transcription (Hanaoka, et al., 2005). Recombinant RpoTp specifically bound to the tRNA molecule in gel mobility shift experiments (Hanaoka, et al., 2005). Moreover, transcription from the plastidial *accD* promoter, which is considered to be catalyzed by a phage-type RNA polymerase, was shown to be inhibited by the addition of $tRNA^{Glu}$ to *in vitro* transcription reactions using proplastid extracts from Arabidopsis as source of transcription activity (Hanaoka, et al., 2005). The authors suggested $tRNA^{Glu}$ to mediate a switch in RNA polymerase utilization from nucleus-encoded RNA polymerases to the plastid-encoded bacterial-type enzyme during chloroplast development.

### I.3.4 Regulation of mitochondrial gene expression at the transcriptional level

Coordinated expression of the nuclear and mitochondrial genomes in individual cells and tissues is required for the assembly of functional mitochondria and ensures appropriate metabolic activities of the organelle in response to environmental stimuli. Substantial progress has been made towards understanding nuclear-mitochondrial interaction in yeast (reviewed in Poyton and McEwen, 1996) and illuminating the signalling pathways between the nucleus and the chloroplast (reviewed in Gray, et al., 2003; Leon, et al., 1998; Pfannschmidt and Liere, 2005; Rodermel, 2001). Besides, models have been proposed that describe transcriptional changes occurring in the chloroplast during organelle biogenesis (Cahoon, et al., 2004; Hanaoka, et al., 2005). In higher plants, retrograde signals between the mitochondrion and the nucleus are established (Millar, et al., 2004); yet only a limited number of studies have addressed the question at what levels mitochondrial gene expression may be controlled.

Transcriptional modulation appears to represent a minor means of regulating gene expression in plant mitochondria (Binder and Brennicke, 2003; Mackenzie and McIntosh, 1999), although tissue-specific differences have been observed in transcript levels for particular mitochondrial genes, e.g. in *in situ* hybridization studies examining various mitochondrial transcripts in maize seedlings (Li, et al., 1996). Similar studies have substantiated a cell-specific regulation of mitochondrial gene expression during sunflower anther development (Smart, et al., 1994). Evidence that such differences are largely the result of posttranscriptional regulation of RNA abundance has been provided by a study that compared transcriptional activities and steady-state RNA levels of mitochondrial genes in

Arabidopsis (Giegé, et al., 2000). While run-on transcription values were found to diverge significantly between genes encoding different subunits of the same protein complex, such differences were less manifest for steady-state RNA levels. Contrary to maize mitochondria where ribosomal RNAs were found to be synthesized at higher rates than other mitochondrial transcripts, thereby accounting for the major contribution of rRNAs to the whole of mitochondrial RNAs (Finnegan and Brown, 1990), the high rRNA levels in Arabidopsis mitochondria were determined to be primarily due to high rRNA stability (Giegé, et al., 2000). In both maize and Arabidopsis, steady-state abundance of protein-coding mRNAs does not correlate with transcriptional processes, emphasizing the importance of posttranscriptional steps in modulating transcript accumulation (Giegé, et al., 2000; Mulligan, et al., 1991). Posttranscriptional processes have also been suggested to be responsible for an observed elevation of mitochondrial transcript levels induced by impaired chloroplast activity in the barley *albostrians* mutant (Hedtke, et al., 1999).

The coordinated expression of the mitochondrial and nuclear genome has been investigated comprehensively in a study employing an Arabidopsis cell culture system to modulate mitochondrial biogenesis in response to sugar starvation and refeeding (Giegé, et al., 2005). A comparison of transcript and protein changes during modulation of sugar supply revealed the mitochondrial genome to be insensitive to sugar starvation stress, whereas changes were observed in the expression of nuclear genes encoding mitochondrial components. Coordination of the expression of mitochondrial and nuclear genes was found to occur at the protein level, possibly during protein-complex assembly (Giegé, et al., 2005).

Limited data support a regulation of mitochondrial gene expression at the transcriptional level. Promoter selection has been reported to be controlled by nucleus-encoded factors in alloplasmic lines of *Nicotiana* and maize (Edqvist and Bergman, 2002; Newton, et al., 1995); these factors and their possible roles in regulating mitochondrial function remain to be identified.

## I.4 Aims of this study

Although it is generally accepted that *RpoT* gene products represent catalytic subunits of the mitochondrial transcription machinery in plants, direct evidence is as yet lacking for RpoT enzymes being involved in the transcription of mitochondrial genes in photosynthetic eukaryotes. Attempts to isolate mitochondrial RNA polymerases, including accessory factors, from plants failed to result in the identification of these components.

The present study aims at reconstituting mitochondrial transcription *in vitro* from recombinant RpoT enzymes, thereby establishing a role of these RNA polymerases in mitochondrial transcription. The nuclear genes *RpoTm* and *RpoTmp* encoding phage-type RNA polymerases that are imported into mitochondria have previously been identified in Arabidopsis. Distinct mitochondrial functions of RpoTm and RpoTmp, the latter of which is also imported into plastids, have so far not been specified. Recombinant Arabidopsis RpoTm and RpoTmp will therefore be examined for possible differences in their transcriptional performances *in vitro*.

Specific transcription initiation by RpoTm or RpoTmp at mitochondrial promoters is likely to require auxiliary factors, which are as yet unknown. Availability of the complete genome sequence of Arabidopsis renders this plant an excellent object of *in silico* analyses directed at identifying candidate transcription factors, based on their similarity to known essential mitochondrial transcription factors in yeast and animals. The subcellular localization and functional properties of these Arabidopsis proteins will be analyzed and compared to those of related yeast and animal factors. Most importantly, *in vitro* transcription experiments will attempt to elucidate if these proteins likewise act as auxiliary factors in mitochondrial transcription.

A prerequisite for studying the mitochondrial transcription machinery is the knowledge of mitochondrial promoters. Sequence motifs identified as elements of mitochondrial promoters in various dicotyledonous species are seen upstream of the coding regions of several mitochondrial genes in Arabidopsis; yet experimental evidence is limited to a single promoter. To define *cis*-elements that direct transcription of the Arabidopsis mitochondrial genome, transcription initiation sites will be mapped and their surrounding sequences, which comprise the promoter, will be analyzed. Experimentally determined promoters will provide a variety of templates for the Arabidopsis *in vitro* transcription system.

# II MATERIALS AND METHODS

## II.1 Growth of Arabidopsis thaliana

*Arabidopsis thaliana* (ecotype Columbia) was grown as described in (Emanuel, et al., 2005).

## II.2 Strains and culturing of *Escherichia coli*

Recombinant plasmids were propagated in *E. coli* Top10 (Invitrogen); cells were grown in LB medium or on LB agar under standard conditions (Sambrook and Russell, 2001). Culturing of *E. coli* for recombinant protein expression is described elsewhere (II.5.2).

## II.3 Nucleic acids

### II.3.1 Isolation of nucleic acids

#### II.3.1.1 Isolation of genomic DNA from Arabidopsis

Genomic DNA was extracted from Arabidopsis leaf tissue using the CTAB method (Murray and Thompson, 1980).

#### II.3.1.2 Plasmid isolation from *E. coli*

Small-scale plasmid preparations were done following the alkaline lysis plasmid miniprep protocol (Sambrook and Russell, 2001); pPROTet.E derivatives were purified via QIAquick spin columns (QIAGEN) following plasmid isolation. Larger amounts of plasmid DNA were isolated from 100-ml cultures using the QIAGEN Plasmid Midi Kit (QIAGEN).

#### II.3.1.3 Isolation of total RNA and mRNA-enriched RNA from Arabidopsis

Total cellular RNA was extracted from leaves and flowers of Arabidopsis plants using TRIZOL (Invitrogen) according to the protocol provided by the manufacturer. The resulting RNA pellet was washed with 70% (v/v) ethanol, resuspended in ultrapure water and column-purified and DNase-treated using the NucleoSpin RNA plant kit (Macherey-Nagel).

Arabidopsis mRNA was enriched from total leaf RNA using the Poly(A)Purist™ Kit (Ambion).

### II.3.2 Determination of nucleic acid concentrations

The quality and quantity of nucleic acids were examined optically in ethidium bromide-stained agarose gels (see II.3.3.1 and II.3.3.2). Additionally, UV absorption at 260 nm by nucleic acid samples was measured in a GeneQuant II photometer (Amersham Biosciences), and concentrations were calculated assuming an optical density $OD_{260}$=1 to correspond to 50 µg/ml double-stranded DNA or 40 µg/ml RNA.

### II.3.3 Nucleid acid electrophoreses

#### II.3.3.1 Agarose gel electrophoresis of DNA

DNA samples of 0.5 to 5 kbp in DNA loading buffer were separated on agarose gels containing 0.8–1.5% (w/v) agarose (Biozym) and 0.2 µg/ml ethidium bromide in 1x TAE buffer. For electrophoreses of 5'-RACE products, gels containing 1% agarose and 2% Nusieve agarose (Biozym) were prepared. Using 1x TAE as running buffer, electrophoreses were carried out at 5-10 V/cm in a horizontal electrophoresis chamber (PerfectBlue Gelsystem Mini S or Mini L, peqlab Biotechnologie GmbH). Lambda DNA (Fermentas GmbH) digested with *Bst*EII or a 1 kb DNA ladder (Invitrogen)

were used as molecular weight markers. Following separation, DNA molecules were visualized by UV transillumination employing a Gel Doc XR System (Bio-Rad).

DNA molecules subjected to preparative agarose gel electrophoreses and excised from gels were purified over QIAquick spin columns (QIAGEN).

| | |
|---|---|
| 1x TAE: | 40 mM Tris; 20 mM acetic acid; 1 mM EDTA |
| DNA loading buffer: | 50% (v/v) glycerol; 1 mM EDTA; 0.005% (w/v) bromphenol blue; 0.005% (w/v) xylene cyanol |

### II.3.3.2 Agarose gel electrophoresis of RNA

1% (w/v) agarose gels for RNA analysis were prepared by melting agarose (Biozym) in ultrapure water supplemented with 10x MEN (1/10 of the final gel volume), and adding formaldehyde (1/40 of the final gel volume) after cooling the matrix to 60°C. RNA samples were supplemented with 1.6 volumes RNA loading buffer and incubated at 65°C for 5 min prior to loading. Using 1x MEN as running buffer, electrophoresis was carried out at 8 V/cm in a horizontal electrophoresis chamber (PerfectBlue Gelsystem Mini S, peqlab Biotechnologie GmbH). Separated RNA molecules were visualized by UV transillumination employing a Gel Doc XR System (Bio-Rad).

| | |
|---|---|
| 1x MEN: | 20 mM MOPS; 5 mM sodium acetate; 1 mM EDTA |
| RNA loading buffer: | 500 µl formamide; 175 ml formaldehyde; 100 µl 10x MEN; 200 µl glycerol, 2.5 µl 0.5 M EDTA, pH 8.0; 5 µl ethidium bromide (10 mg/ml); 2 mg bromphenol blue; 2 mg xylene cyanol; ultrapure water ad 1 ml |

### II.3.3.3 Denaturing polyacrylamide gel eletrophoresis (PAGE) of RNA

Denaturing PAGE was made use of to resolve products of *in vitro* capping analyses and *in vitro*-synthesized RNAs. Separation in 0.75-mm-thick 5% acrylamide gels was carried out utilizing a Protean II xi electrophoresis unit (Bio-Rad); a Model S2 Sequencing Gel Electrophoresis Apparatus (Biometra GmbH) was employed for high-resolution electrophoresis of *in vitro*-synthesized RNAs in 0.4-mm-thick 5% acrylamide sequencing gels. A radiolabelled RNA length standard was generated using the RNA Century Marker Template Plus (Ambion) and MAXIscript kit (Ambion) according to the manufacturer's instructions and separated alongside RNA samples. Using 0.6x TBE as electrophoresis buffer, gels were run at 25 mA per gel (Protean II xi) or at 55 W (sequencing gels). Following a 10-min pre-run of gels, RNA samples dissolved in formamide buffer and denatured at 95°C for 5 min were loaded and electrophoresed for 2 h. Gels were then transferred on Whatman 3MM paper, dried on a Model 583 Gel Dryer (Bio-Rad) and subjected to autoradiography employing a phosphoimager (Molecular Imager FX, Bio-Rad).

| | |
|---|---|
| 1x TBE: | 90 mM Tris; 90 mM boric acid; 1 mM EDTA |
| Acrylamide stock solution: | Gel 40 (Roth) |
| Gel composition: | 7 M urea and 5% acrylamide in 1x TBE |

### II.3.3.4 Native PAGE of DNA

DNA and DNA/protein complexes were resolved by native PAGE in 5% polyacrylamide gels at 4°C employing a Protean II xi electrophoresis unit (Bio-Rad) at 4°C; 0.5x TBE (see II.3.3.3) was used as electrophoresis buffer. Gels were pre-run at 200 V for 1-1.5 h prior to sample loading. DNA and DNA/protein complexes were separated at 15 mA per gel for 4-6 h; a dye marker was alongside samples to monitor migration. Gels were transferred on Whatman 3MM paper, dried on a Model 583 Gel Dryer (Bio-Rad) and subjected to autoradiography employing a phosphoimager (Molecular Imager FX, Bio-Rad).

| | |
|---|---|
| Acrylamide stock solution: | Gel 30 (Roth) |
| Gel composition: | 5% acrylamide in 0.5x TBE |
| Dye marker: | 0.6 M Tris/HCl, pH 6.8; 50% (v/v) glycerol; 0.4% (w/v) bromphenol blue |

### II.3.4 cDNA synthesis and RT-PCR

cDNA was made from 1 µg mRNA-enriched Arabidopsis RNA employing the Omniscript RT kit (QIAGEN, Germany) according to the manufacturer's instructions; 250 nmol of a random hexamer primer mixture (Fermentas) were used to prime first strand cDNA synthesis. Reactions were allowed to proceed at 42°C for 2 h.

Protein-coding sequences were amplified from Arabidopsis cDNA using *Pfu* DNA polymerase (Promega). PCR reactions were carried out in a volume of 50 µl in the appropriate buffer (Promega) with 2.5 U DNA polymerase, 10 pmol each of the forward and reverse primer, 10 µmol of each dNTP, and 0.5 µl of the cDNA synthesis reaction. Cycling conditions were as follows: 94°C/1 min; 35 cycles of 95°C/20 s, 58-62°C/20 s, 72°C/2 min per 1 kbp of amplicon length; 72°C/10 min. PCR was carried out in a Peltier Thermal Cycler PTC-200 (Biozym). PCR Products were analysed by agarose gel electrophoresis (II.3.3.1).

### II.3.5 PCR

DNA fragments were amplified from genomic Arabidopsis DNA using *Taq* DNA polymerase (QIAGEN). Reactions contained 1 U DNA polymerase, 10 pmol each of the forward and reverse primer, 10 µmol of each dNTP, and 50 ng genomic DNA in 50 µl of the appropriate buffer. Cycling conditions were as follows: 94°C/1 min; 35 cycles of 95°C/20 s, 58-62°C/20 s, 72°C/1 min per 1 kbp of amplicon length; 72°C/6 min.

For colony PCR analysis of cloned DNA fragments, reactions were set up with 0.5 U *Taq* DNA polymerase (QIAGEN), 5 pmol each of the forward and reverse primer, 5 µmol of each dNTP and cells from a bacterial colony in 25 µl of the appropriate buffer; cycling was done as described above.

PCR Products were analyses by agarose gel electrophoresis (II.3.3.1).

### II.3.6 Cloning and sequencing

Standard procedures for DNA manipulation such as restriction digests and ligations of DNA molecules were carried out according to (Sambrook and Russell, 2001). Restriction endonucleases, Shrimp Alkaline phosphatase and T4 DNA Ligase were obtained from Fermentas. The QIAGEN PCR Cloning Kit was used to directly ligate PCR products into the pDrive vector (QIAGEN).

#### II.3.6.1 Transformation of *E. coli*

Plasmid DNA was introduced into *E. coli* by electroporation (Dower, et al., 1988) using a Gene Pulser electroporation device (Bio-Rad). Transformants were selected on solid LB medium containing the appropriate antibiotics. If applicable, LB plates were supplemented with X-Gal for blue/white selection transformed TOP10 cells (Sambrook and Russell, 2001).

#### II.3.6.2 Sequencing

Sequencing reactions were set up using the ABI PRISM™ Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) according to the manufacturer's instructions. Reactions contained 300-400 ng plasmid DNA or 15-40 ng PCR fragment and 5-10 pmol primer (see Table 1 for oligonucleotides used as sequencing primers). Cycle sequencing, product purification and product analysis on an ABI 377 automatic DNA Sequencer (Applied Biosystems) were carried out by M. Meixner (DLMBC, Rüdersdorf).

**Table 1: Oligonucleotides used for sequencing.**

| Primer | Sequence (5'→3') | Target |
|---|---|---|
| metA-seq1m | CAACCACGAGATCAAAGTG | *MetA* |
| metA-seq2p | AGCTAGATAAGCGTATGGTGG | |
| metA-seq3p | CTTGGTTCCATGTTTAGGC | |
| metB-seq1m | TCCTGACAACGGAAGAATC | *MetB* |
| metB-seq2p | TCTACTGCCGTCTCTCTGTC | |
| RpoT1-522R | AACCAGCCTAGAAACAAAGAC | *RpoTm* |
| Y2 | GTCGCATTGGTGGTCTGG | |
| Y6 | AGGTCGGATGTCAGGTGG | |
| Y11 | CAGAAGCCTTGAGAAGCCCC | |
| Y15 | AATTGCTTTTTTGGATTCCC | |
| RpoT1-2400F | AAGCTGCAAGAGCTATCAAG | |
| T2-621R | GCAGATTAGGCGCAAGC | *RpoTmp* |
| T2-850R | GGACATCAGGCAGATCATT | |
| 3cF1 | CAACAGATGTTGAGGAAGAGCC | |
| 3cF2 | AAAAGGGGATGACAATGAGG | |
| 3cR1 | CATTCACCAAACCAACGC | |
| 3cR2 | CAGCCACCATCTGCTTCC | |
| RpoT2-2560F | TTTGGTGAATGTGCGAAG | |
| M13-seq-F | ACGACGTTGTAAAACGACGG | pDrive, pKL23 |
| M13-seq-R | TTCACACAGGAAACAGCTATGAC | |
| pPROTet-seqF | TCATTAAAGAGGAGAAAGGTACCC | pPROTet.E |
| pPROTet-seqR | CCATGGGTACCTTTCTCCTCT | |
| pBAD-rev | GATTTAATCTGTATCAGG | pBAD/Thio-TOPO |
| gfp-seq3 | GCCAAGGAACAGGTAGTT | pOL−GFP S65C |

### II.3.7 5'-RACE analysis of RNA

Mitochondrial transcript 5' termini were determined employing a 5'-RACE technique described by (Bensing, et al., 1996) with the following modifications. 5' triphosphates were converted to monophosphates by treating 5 µg RNA with 10 U of tobacco acid pyrophosphatase (Epicentre) at 37°C for 1 h in the presence of 40 U of RNase inhibitor (Fermentas) in the appropriate buffer. Control reactions were set up without pyrophosphatase. The RNA was subsequently extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated from the aqueous phase by adding 3 volumes of ethanol/3 M sodium acetate, pH 5.2 (30:1) and dissolved in ultrapure water. The RNA was then supplemented with 10 pmol 5' RNA adapter A3 [5'-GAUAUGCGCGAAUUCCUGUAGAACGAACACUAGAAGAAA-3', (Argaman, et al., 2001)], and the ligation of transcripts to the adapter was performed at 37°C for 1 h with 50 U of T4 RNA ligase (Epicentre Technologies) in the presence of 1 mM ATP and 80 U of RNase inhibitor (Fermentas) in the appropriate buffer. Control reactions were set up without adding the adapter. Following the ligation, the RNA was extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated by adding 3 volumes of ethanol/3 M sodium acetate, pH 5.2 (30:1), dissolved in ultrapure water and then reverse-transcribed using gene-specific primers and Omniscript RT kit (QIAGEN) according to the protocol provided by the manufacturer; reactions were allowed to proceed for 2 h at 42°C. The products of reverse transcription were amplified in a first PCR step by using 1 to 3 µl of the RT reaction, 5 pmol of each adapter-specific forward primer P1a (5'-CGAATTCCTGTAGAACGAACACTAGAAG-3') and gene-specific reverse primer, 200 µM of each dNTP and 0.5 U of *Taq* DNA polymerase (QIAGEN) in 25 µl of the appropriate buffer. Cycling conditions: 94°C/1 min; 35 cycles of 95°C/20 s, 58-62°C/20 s, 72°C/2 min; 72°C/10 min. 0.1 to 1 µl of the first PCR reaction were used as template for subsequent nested PCRs set up essentially as the first PCR in a volume of 50 µl with 10 pmol of each gene-specific and adapter-specific primer. Gene-specific primers were repeatedly placed upstream of identified transcriptional starts, until no 5'-RACE products reaching further upstream could be detected. PCR reactions were analysed by agarose gel electrophoresis. Products of interest were excised and ligated into pDrive (QIAGEN); ligation products were transformed into *E. coli* TOP10 (Invitrogen). Bacterial clones containing the plasmid

insert were identified by colony PCR using primers M13-seq-F and M13-seq-R. Colony PCR reactions were set up and performed essentially as the first PCR of the 5'-RACE protocol, and PCR products were purified over QIAquick spin columns (QIAGEN) and sequenced.

Oligonucleotides listed in Table 2 were used as reverse primers in 5'-RACE experiments for reverse transcription (application "A"), the first RACE PCR ("B") or subsequent nested PCRs ("C").

**Table 2: mtDNA-specific primers used for 5'-RACE.**

| Primer | Gene/ RNA | Sequence (5'→3') | Application | Detected start site |
|---|---|---|---|---|
| P2rrn18 | *rrn18* | CGAGAACAACGTTCGAC | A | |
| P3rrn18-b | | GCCCTGCAGTGGTAGAACCTC | B | |
| P4rrn18-b | | TCGTGAACCGGGCGTACTAC | C | P*rrn18*-69<br>P*rrn18*-156 |
| P5rrn18 | | TTCTATCAATCGATAAGCAAGGGTAGG | C | P*rrn18*-353<br>P*rrn18*-424 |
| P7rrn18 | | TTATATACCGAGGATTTGATGAAATACCA | C | - |
| P2rrn26 | *rrn26* | AGTGCGCTTGACTACTCCT | A | |
| P3rrn26 | | TGATTGTTCGCATCGGATCTC | B | |
| P4rrn26-b | | TCACGGTACTTGTACGCTATCGG | C | - |
| P5rrn26 | | CCGCTCACAGAAGGATTTACAGTC | C | P*rrn26*-893 |
| P6rrn26 | | GAGGGATGCAATAACTCGACTGTG | C | - |
| P7rrn26 | | TATCCGAAACATTTTTGAACTGCC | C | - |
| P2met | *tRNA-fMet* | CCGCTTCTTCTCTTCTACAAG | A | |
| P3met | | TTCTAGAGACAAACGACCGATTGAA | B | |
| P4met | | AGACAAACGACCGATTGAACTACAA | C | P*trnM*-98 |
| P5met | | TCAAAAGAAAGAAGTAGAGTCGTTGGAC | C | P*trnM*-547 |
| P7met | | GTGGAAACAACTCCCTTAGCCTTAG | C | - |
| P2rps3 | *rps3* | GCCCTCACTGAACCGACT | A | |
| P3rps3 | | GAACCGACTTGAATCTGAACTACGA | B | |
| P4rps3 | | TTTTTGACTTTATGGATTTCTGTCCCT | C | - |
| P7rps3 | | AGATAGAAATGATAGAGGGCCAACC | C | P*rps3*-1053<br>P*rps3*-1133 |
| P8rps3 | | AGCTAACGTAAGAACTGGAAGAGTCTTG | C | - |
| P2atp1 | *atp1* | TATATGGATTCGGGCTGC | A | |
| P3atp1 | | GAAGTAGCGCGAGAAGGTACGA | B | |
| P4atp1 | | CGATACCAGTTGGGCGAACA | C | - |
| P5atp1 | | AGTAGACGGAACGACACCTGTGA | C | P*atp1*-1898 |
| P6atp1 | | GGCTACTTTCTTTCTTCTCTTATGAAATTG | C | P*atp1*-1947 |
| P7atp1 | | ATACCACCAGATGTGCCCCTT | C | - |
| P8atp1 | | TCCTTTTCTTTTTGAGCAGATGTTG | C | - |
| P2atp6-1 | *atp6-1* | GGGATCTTGCGTTAATGC | A | |
| P3atp6-1 | | GATCTTGCGTTAATGCCTCACAC | B | |
| P4atp6-1b | | CAAACAAAAAGATTCGTCGCATATTG | C | P*atp6-1*-156<br>P*atp6-1*-200 |
| P6atp6-1 | | GATTTGGAAGGGCAAGATAGACC | C | - |
| P7atp6-1 | | CGGTTCATCGCCTTACTTATCCA | C | P*atp6-1*-916/913 |
| P8atp6-1 | | CCTAATCAAGCAGAACGCCACT | C | - |
| P9atp6-1 | | GCCCTCAGCAGCTCGAATAC | C | - |
| P2atp6-2 | *atp6-2* | AAGTGATTCAACCGGGTTA | A | |
| P3atp6-2 | | GAATAGGCACTCCTGGCAGAAC | B | |
| P4atp6-2 | | AGATTTGGCTTTTGAGGCATGA | C | P*atp6-2*-148<br>P*atp6-2*-436<br>P*atp6-2*-507 |
| P6atp6-2 | | GAGTAGCAAAGATGACAGCACGC | C | - |
| P6atp6-2b | | CGCACAAACATATCCGACTCGTA | C | - |
| P2atp8 | *orfB* | AAACTGTTGGGGTCCTTG | A | |
| P3atp8 | | TGTGAAAGCAGTTGGTTCCGTAG | B | |

| | | | | |
|---|---|---|---|---|
| P2atp8 | | AAACTGTTGGGGTCCTTG | A | |
| P3atp8 | | TGTGAAAGCAGTTGGTTCCGTAG | B | |
| P4atp8 | | AGAAAGTAAAGAAGAAAAGGCATAACCAG | C | Patp8-157<br>Patp8-228 |
| P6atp8 | | TCGTTAGAAGAAGATGAGCTGCCT | C | - |
| P7atp8 | | CAGCCCGGATCACCAGCTA | C | Patp8-710 |
| P8atp8 | | ACCAGCAGAATTAGATGAACGAGC | C | Patp8-999 |
| P9atp8 | | GGGTCTTTAGAGGACTATGCCAAGT | C | - |
| P2atp9-c | *atp9* | TGCAATAGCTTCGGTTAGAG | A | |
| P3atp9-b | | TTTAGCCAATGATGGATTTCGC | B | |
| P4atp9-a | | GACTGAAGACCAACTGAATCTCGAC | C | P*atp9*-239<br>P*atp9*-295 |
| P5atp9 | | TTATATACGAGAGCACCAGATACACCA | C | P*atp9*-487<br>P*atp9*-652 |
| P7atp9 | | GAAGATCAAGTTACTCGGCTAGACCA | C | - |
| P2cox1 | *cox1* | TGTGCCCATCACTCCAG | A | |
| P3cox1 | | ACCGAAAATGAAATAGAGAGTCCCT | B | |
| P4cox1 | | TGTGGTTTGTGGAGAACAGCC | C | - |
| P5cox1-b | | ATCGTCCTACAAAAGATAATGCTCTCAC | C | P*cox1*-355 |
| P6cox1 | | GCCACATTTCATACACTTTAGGCA | C | - |
| P7cox1 | | CCAGCAGCTACAACCAAGTCAG | C | - |
| P2cox2-c | *cox2* | CCGAAGAATCTCGATAGTAG | A | |
| P3cox2 | | GTAGCTGCGTCTTGAGATCCTAATTG | B | |
| P3cox2-d | | TTCTTCTTCTTCTTACAATATTTTGAGTTAGATG | C | P*cox2*-210 |
| P5cox2-d | | CGAAACCAACATCCTTATAATACTACTAGGC | C | P*cox2*-481 |
| P7cox2 | | CATTAGATAGCTAATTATCCTTTGCCTAGC | C | P*cox2*-683 |
| P8cox2 | | GGTAGGGCTCTGTTTCAGGTCTTG | C | - |
| P9cox2 | | ATGGCTGGTTGAGGTTAGAATTTC | C | - |
| P2cox2-c | *orf291* | CCGAAGAATCTCGATAGTAG | A | |
| P3cox2 | | GTAGCTGCGTCTTGAGATCCTAATTG | B | |
| P4cox2 | | CAAGGAGAAATTGTGAGGAATAACCA | C | P*orf291*-307 |
| P2nad1-int-a | *nad1*-AS | AGTTGCGATGCGAACAG | A | - |
| P3nad1-int-a | | GAGTAGACTTGCCTGAGTTGTCTGC | B | - |
| P4nad1-int-a | | TTCATTTTCTTTTAGTTGCGGTAGC | C | P1*nad1*-AS |
| P2nad1-int-b | | GCATCGCGATAAGTCCTC | A | - |
| P3nad1-int-b | | GCAATATTCACCCTAGCCCACAA | B | - |
| P4nad1-int-b | | GCCGAATATAATCCTCAAGTACTCCA | C | P2*nad1*-AS |
| P2nad4-int | *nad4*-AS | TTGTAGGTGCTTGCGATG | A | - |
| P3nad4-int | | AGTTGGTTTGGGTGGCATAGC | B | - |
| P4nad4-int | | TAGCCCGTTGCATAAGTCCC | C | P1*nad4*-AS<br>P2*nad4*-AS |
| P2nad5-int | *nad5*-AS | CCCGACTCTACGAACCC | A | - |
| P3nad5-int | | CCCGAGGAAAGGCTGCAC | B | - |
| P4nad5-int | | CAGTAGTAAGGGCGTTAAGACCGA | C | P*nad5*-AS |
| P2nad2-int | *nad2*-AS | TTTGTATTATAAGTGATCCGAACC | A | - |
| P3nad2-int | | CCAAGTTGGTGAGCCGTATGAT | B | - |
| P4nad2-int | | CGGTTTGGAGAGGACTCAGC | C | P*nad2*-AS |
| P2nad7-int | *nad7*-AS | CTTTGCCGAGAGATAGGAG | A | - |
| P3nad7-int | | CTTTGAGAACTGTGTGAACGGAGAG | B | - |
| P4nad7-int | | GAATGGGTCGAGATAGATGACAGC | C | P*nad7*-AS |
| P2mot3-38K | NC | CGAGGGTTCAATTCAGTG | A | - |
| P3mot3-38K | | AATAGCTAGATACTCTGCGGGACCTC | B | - |
| P4mot3-38K | | TCTTTCTAATTAATCGTTTTACCGGGAATAC | C | $P_{38K}$-nc |
| P2-203K | NC | GGTGCTTTCAGGAACTGG | A | - |
| P3-203K | | GGAGACGGGCTATGTAGGCTG | B | - |
| P4mot2-203K | | AAAGAAGGAAAGGATAGTATTCGGTGG | C | $P_{203K}$-nc |

### II.3.8 Analysis of *in vitro*-cappable transcripts

#### II.3.8.1 Preparation of riboprobes

Sequences of upstream regions of the mitochondrial genes *atp1*, *atp6-1*, *atp6-2*, *atp9*, *cox2*, *rrn18* and *rrn26* were amplified from total Arabidopsis DNA with primer pairs listed in Table 3, and ligated into pDrive (QIAGEN) in the appropriate orientation to yield templates for complementary RNA (cRNA) synthesis (Table 3). Riboprobes were generated through *in vitro* transcription of antisense strands of the cloned gene fragments by T7 RNA polymerase and a subsequent DNase digest using the MAXIscript kit (Ambion). Transcripts were then extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated from the aqueous phase by adding 3 volumes of ethanol/3 M sodium acetate, pH 5.2 (30:1), and dissolved in ultrapure water. See Figure 12 for target sequences of riboprobes.

**Table 3: Primers used for cRNA synthesis template construction.**

| Primer pair | Primer sequences (5'→3') | cRNA synthesis template |
|---|---|---|
| Cap-atp1-F<br>Cap-atp1-R | GCTAGGCTGGCACTTAGGA<br>TGATAGGTTTATTTCGCACTTTAG | pCRNA-atp1 |
| Cap-atp6-1-F<br>Cap-atp6-1-R | CACCGCACGAGTCAGACCT<br>TCCAGACAGCTTCACTCCGTC | pCRNA-atp6-1 |
| Cap-atp6-2-Fb<br>Cap-atp6-2-Rb | CGTGTTCCGTCGATCAC<br>CTTACGTCAAGCCTCTAGGAGT | pCRNA-atp6-2b |
| Cap-atp9-Fb<br>Cap-atp9-Rb | TTGGGATAAGTGAAATCGTAT<br>CGACAAAGAGAAGTACAAGC | pCRNA-atp9-b |
| Cap-cox2-F<br>P3cox-d | TGCCTTGCCTTACCACACC<br>see Table 2 | pCRNA-cox2 |
| Cap-rrn18-F<br>P4rrn18-b | GAGACCGATCCAGGAACCCTAC<br>see Table 2 | pCRNA-rrn18 |
| Cap-rrn26-F<br>Cap-rrn26-R | AAAGGCGTTATTGCTGTGCT<br>TTTTCAACTCGTAAAGGCAAAGA | pCRNA-rrn26 |

#### II.3.8.2 *In vitro* capping and RNase protection

*In vitro* capping reactions were set up in a volume of 20 µl with 20 µg of total RNA isolated from flowers and 5 U guanylyltransferase (Ambion) in the appropriate buffer in the presence of 130 µM S-adenosyl methionine, 2.5 U RNase inhibitor (Fermentas), 100 µCi [$\alpha$-$^{32}$P]-GTP (3000 Ci/mmol), and were incubated at 37°C for 75 min. After 30 min, another 7.5 U of guanylyltransferase were added. The RNA was purified with 2 volumes of ultrapure water, 4 volumes of TRIzol (Invitrogen) and 0.8 volumes of chloroform, precipitated from the aqueous phase by adding 0.8 volumes of isopropanol and washed twice with 70% ethanol. Transcripts were then dissolved in 30 µl hybridization buffer (Roche) together with 0.5 µg of complementary riboprobe and subjected to ribonuclease protection using the RNase Protection Kit (Roche) according to the protocol provided by the manufacturer, except that hybridizations were carried out overnight at 45°C (*rrn18*, *rrn26*) or 65°C (all other genes). Protected transcripts were dissolved in formamide loading buffer provided with the kit and separated in 5% polyacrylamide gels (II.3.3.3).

## II.4 Protein analysis

### II.4.1 Determination of protein concentrations

Protein concentrations of bacterial lysates and lysate fractions were compared to a BSA standard as described by Bradford (1976) using the Bio-Rad Protein Assay. Concentrations of distinct proteins were approximated by comparison to defined BSA amounts in Coomassie-stained polyacrylamide gels.

### II.4.2 SDS polyacrylamide gel electrophoresis (SDS PAGE)

For total protein analysis, cells pelleted from 200 µl aliquots of *E. coli* cultures were lysed in 40 µl 1x sample buffer at 95°C for 5min. Following centrifugation, an appropriate aliquot of the supernatant was subjected to SDS PAGE. Bacterial lysates and lysate fractions in 1x sample buffer were similarly prepared.

Protein separation by SDS PAGE was carried out according to Laemmli (1970) in a Hoefer Mighty Small Vertical Electrophoresis Unit. Electrophoresis was allowed to proceed at 200 V for 1-1.5 h; a protein molecular weight marker (#0661 or #SM671, Fermentas) was run alongside samples. Following electrophoresis, gels were Coomassie-stained or subjected to Western blotting.

| | |
|---|---|
| 4x sample buffer: | 0.32 M Tris/HCl pH 6.8, 0.1 M EDTA, 0.4 M DTT, 8% (w/v) SDS, 4% (v/v) glycerol, 0.2 % (w/v) bromphenol blue |
| Acrylamide stock solution: | Gel 30 (Roth) |
| Separating gel: | 8% or 10% acrylamide, 375 mM Tris/HCl pH 8.8, 0.1% (w/v) SDS |
| Stacking gel: | 4% acrylamide, 125 mM Tris/HCl pH 6.8, 0.1% (w/v) SDS |
| Electrophoresis buffer: | 25 mM Tris, 192 mM glycin, 0.1% (w/v) SDS |
| Coomassie staining solution | 0.1% (w/v) Coomassie Brilliant Blue R250 in destaining solution |
| Destaining solution: | 0.25 % (v/v) isopropanol, 0.1 % (v/v) acetic acid |

### II.4.3 Immunoblotting

Protein patterns resolved by SDS-PAGE were transferred onto nitrocellulose membranes (Hybond-C extra, Amersham Biosciences) by electrotransfer in a Semi Dry Blot chamber (Bio-Rad). Following PAGE, gels were incubated in transfer buffer for 10 min. Whatman 3MM paper and the membrane were preincubated in transfer buffer. The anode plate was overlaid with 4 layers of 3MM paper, followed by the membrane, the gel, and 4 layers of 3MM paper. Electrotransfer was allowed to proceed at 20 V for 1 h.

Membranes were subsequently blocked for 2 h in TBS containing 3% (w/v) BSA, washed 4x 10 min in TBST, incubated with the primary antibody diluted in TBST (Table 4) for 1.5 h, washed 4x 10 min in TBST, incubated with the secondary antibody diluted in TBST (Table 4) for 1 h, washed 4x 10 min in TBST and then 10 min in TBS, and rinsed 3x 1 min in AP buffer. Immunodetection using alkaline phosphatase was initiated by exposure of membranes to AP sustrate solution; appropriately developed membranes were washed in water to stop the enzymatic reaction.

| | |
|---|---|
| Transfer buffer: | 48 mM Tris, 39 mM glycin, 20% (v/v) methanol, 0.0375% (w/v) SDS |
| TBS: | 10 mM Tris/HCl pH 7.5, 0.2% (w/v) SDS |
| TBST-BSA: | 1% (w/v) BSA in TBST |
| TBST: | 0.05 % (v/v) Tween 20 in TBS |
| AP buffer: | 100 mM Tris/HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$ |
| AP substrate solution | 0.033% (w/v) NBT, 0.0165% (w/v) BCIP in AP buffer |

**Table 4: Antisera.**

| Antibody | Properties | Dilution | Reference |
|---|---|---|---|
| *anti*-Thio | Mouse IgG, raised against thioredoxin | 1:5000 | Invitrogen |
| *anti*-His | Mouse IgG, raised against a polyhistidine-tagged protein | 1:5000 | Sigma |
| *anti*-RpoT | Rabbit IgG, raised against the C-terminal portion (amino acids 663-976) of RpoTm from Arabidopsis | 1:3000 | A. Weihe, HU Berlin |
| Secondary antibody | Anti-rabbit IgG-alkaline phosphatase conjugate | 1:5000 | Sigma |
| Secondary antibody | Anti-mouse IgG-alkaline phosphatase conjugate | 1:5000 | Sigma |

## II.5 Recombinant protein expression

### II.5.1 Plasmids for the expression of recombinant proteins

Sequences encoding amino acids 43-976 of RpoTm (locus tag At1g68990) and amino acids 105-1011 of RpoTmp (locus tag At5g15700) were amplified from reverse-transcribed mRNA-enriched Arabidopsis RNA using primer pairs His-RpoTm-42P/pBAD-Yrev and His-RpoTmp-104P/pBAD-3rev, respectively. PCR products were ligated into pBAD/Thio-TOPO (Invitrogen) to yield plasmids pBAD/Thio-HisRpoTm and pBAD/Thio-HisRpoTmp encoding thioredoxin (Trx)-hexahistidine-tagged RpoTm and RpoTmp.

Additional constructs were generated for the expression of untagged RpoTm and RpoTmp. Equimolar amounts of oligonucleotides duet-Linker1-Fw and duet-Linker1-Rev in Taq buffer (QIAGEN) were allowed to anneal at room temperature following incubation at 95°C for 5 min. The double-stranded fragment having Acc651/Pst1 overhangs was used to replace an Acc651/Pst fragment excised from pPROTet.E121 (Clontech). The excised fragment precedes the multicloning site in pPROTet.E121 and encodes a hexahistidine tag. The resulting vector designated pPROL1 directs the expression of essentially untagged recombinant proteins. To generate pPROL1-RpoTm and pPROL1-RpoTmp, the previously cloned RpoTm and RpoTmp coding sequences were amplified from pBAD/Thio-HisRpoTm and pBAD/Thio-HisRpoTmp using primer pairs Duet-T1-F/Duet-T1-R and Duet-T2-104-F/Duet-T2-R, respectively. Following ligation into pDrive and control of PCR products by sequencing, RpoTm and RpoTmp sequences were excised and ligated into pPROL1 via the PstI/PvuI and BamHI/PvuI sites, respectively.

Sequences encoding amino acids 26-380 of MetA (locus tag At5g66360) and amino acids 18-353 of MetB (locus tag At2g47420) were amplified from reverse-transcribed mRNA-enriched Arabidopsis RNA using primer pairs MetA-PROP/MetA-PROM and MetB-PROP/MetB-PROM, respectively. PCR products were NotI/SalI-digested and ligated into the NotI/SalI-cleaved vector pPROTet.E121 to generate plasmids pPRO-MetA and pPRO-MetB.

Plasmids constructed for recombinant protein expression and oligonucleotides employed for plasmid construction are listed in Table 5.

**Table 5: Oligonucleotides used to construct plasmids driving recombinant protein expression.** Lowercase nucleotides correspond to non-annealing sequences added in order to introduce a hexahistidine-encoding sequence or restriction sites.

| Oligonucleotide Pair | Sequence (5'→3') | Plasmid |
|---|---|---|
| His-RpoTm-42P<br>pBAD-Yrev | catcatcatcatcatcatGGCGTTAGAAATGGTTTATCTATAA<br>TGCAGCTCAGTTGAAGAAGTATG | pBAD/Thio-HisRpoTm |
| His-RpoTmp-104P<br>pBAD-3rev | catcatcatcatcatcacGAGTTTTCCAAGAGCGAGAG<br>TCAGTTGAAGAAATAAGGTGAATC | pBAD/Thio-HisRpoTmp |
| duet -Linker1-Fw<br>duet -Linker1-Rev | GTACCCATGGGTGTGGCAGGCGGGGGCGGATCCCTGCA<br>GGGATCCGCCCCCGCCTGCCACACCCATGG | pPROL1 |
| Duet-T1-F<br>Duet-T1-R | gctgcagGGCGTTAGAAATGGTTTATCTATAA<br>gcgatcgGCAGCTCAGTTGAAGAAGTATGT | pPROL1-RpoTm |
| Duet-T2-104-F<br>Duet-T2-R | gggatccGAGTTTTCCAAGAGCGAGAG<br>gcgatcgTCAGTTGAAGAAATAAGGTGAATC | pPROL1-RpoTmp |
| MetA-PROP<br>MetA-PROM | cagcgtcgacCGAGATTCTCACTCGCAGGC<br>cagcgcggccgcTTATTCGTGTAGATCCATTTGTAATGATG | pPRO-MetA |
| MetB-PROP<br>MetB-PROM | cagcgtcgacTCGAACCATTACCAAGGAGGAATAT<br>cagcgcggccgcACACCACAAAACGATTATGTGAAGTG | pPRO-MetB |

### II.5.2 Protein expression in *E. coli*

RpoTm and RpoTmp were overexpressed from pBAD/Thio-HisRpoTm and pBAD/Thio-HisRpoTmp in *E. coli* strain BL21 Codon Plus RIL (Stratagene). 2.5 ml of an overnight culture grown

under standard conditions in LB medium supplemented with 100 µg/ml Ampicillin and 30 µg/ml Chloramphenicol were used to inoculate 250 ml fresh LB medium containing 100 µg/ml Ampicillin. Cultures were grown under standard conditions; recombinant protein expression was induced at $OD_{600}$~0.8 by adding 0.02% (w/v) arabinose, and cells were cultured at 18°C for 20 h until harvest by centrifugation (10 min, 6000xg, 4°C) in a Megafuge 1.0 R (Heraeus).

Strain *E. coli* BL21PRO (Clontech) was employed to overexpress MetA and MetB from pPRO-MetA and pPRO-MetB. 2.5 ml of an overnight culture grown under standard conditions in LB medium containing 34 µg/ml Chloramphenicol and 50 µg/ml Spectinomycin were used to inoculate 250 ml fresh LB medium supplemented with the same antibiotics. Cultures were grown under standard conditions; recombinant protein expression was induced at $OD_{600}$~0.8 by adding 80 ng/ml anhydrotetracycline, and cells were cultured at 18°C for 6 h (MetA expression) or 20 h (MetB) until harvest by centrifugation (10 min, 6000xg, 4°C) in a Megafuge 1.0 R (Heraeus).

Expression of RpoTm and RpoTmp in *E. coli* BL21PRO harbouring pPROL1-RpoTm and pPROL1-RpoTmp, respectively, was monitored in 3-ml cultures over 20 h following induction; the above described protocol was downscaled accordingly.

200-µl aliquots of cultures were pelleted and subjected to SDS-PAGE analysis (II.4.2) of recombinant protein expression.

### II.5.3 Purification of recombinant proteins from *E. coli*

#### II.5.3.1 Trx-$(His)_6$-tagged RpoTm and RpoTmp

Recombinant RpoTm was prepared from 300 ml of cell culture; RpoTmp was prepared from a culture volume of 150 ml. Unless indicated otherwise, centrifugations were done in a in a Megafuge 1.0 R (Heraeus) or in a Biofuge fresco (Heraeus). Harvested cells were resuspended 300 ml (RpoTm) or 150 ml (RpoTmp) buffer A1, recentrifuged and resuspended in 7.5 ml buffer A2 and distributed over 10 2-ml microcentrifuge tubes containing ~1 g glass beads (∅ 0.17-0.18 mm) each. Cells were broken by shaking in a bead mill (Retsch) at maximum frequency (10 min, 4°C). The bacterial lysate was cleared by centrifugation (10 min, 6000xg, 4°C) and transferred into a fresh 15-ml polypropylene tube. Beads were twice washed in ~300 µl buffer A2 per tube by additional grinding (5 min) and centrifugation. Soluble fractions obtained from all extraction steps were pooled to obtain 12-15 ml of bacterial lysate and recentrifuged (10 min, 6000xg, 4°C). The lysate and 0.5 ml (bed volume) $Ni^{2+}$-NTA-agarose (QIAGEN) equilibrated with buffer A1 were distributed over 3 15-ml polypropylene tubes and incubated horizontally while slowly rocking (15 h, 70 rpm, 4°C) to allow proteins to bind to the matrix. Subsequently, unbound proteins were removed by centrifugation (2 min, 1000xg, 4°C) and removal of the supernatant. The matrix was then washed by agitation in 6 ml buffer A3, recentrifuged, resuspended in 3 ml buffer A3, and the slurry was transferred into a 1-ml polypropylene column (QIAGEN). The settled matrix was washed with 3x 2 ml buffer A4. Proteins were then eluted with 3x 1 ml buffer AE. The volume of the eluate was brought to 1 ml by centrifugation (9000xg, 4°C) in 10000-MWCO centrifugal filter devices (Millipore) in a Sorvall RC-5B centrifuge/rotor SS-34 (DuPont) and dialyzed for 15 h at 4°C against buffer AD. Aliquots of dialyzed protein were stored at -20°C for use within 6 weeks. Protein preparations were analysed by SDS-PAGE; a molecular weight of 117 kDa was assumed for recombinant RpoTm and RpoTmp to approximate the molar concentration of proteins in protein preparations.

| | |
|---|---|
| Buffer A1: | 100 mM Tris/HCl, pH 7.8; 300 mM NaCl; 5 mM imidazole |
| Buffer A2: | Buffer A1 supplemented with 1 mM PMSF, 1 mM benzamidine, 0.5 mM DTT |
| Buffer A3: | 20 mM Tris/HCl, pH 7.0; 300 mM NaCl; 5 mM imidazole |
| Buffer A4: | 20 mM Tris/HCl pH 7.0; 300 mM NaCl; 10 mM imidazole |
| Buffer AE: | 20 mM Tris/HCl pH 7.0; 300 mM NaCl; 10 mM imidazole |
| Buffer AD: | 20 mM Tris/HCl, pH 7.8, 100 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 50% (v/v) glycerol |

### II.5.3.2 Proteolytic removal of thioredoxin

Enterokinase digest reactions were allowed to proceed for 15 h at 4°C in a volume of 600 µl containing 1.5 U Enterokinase (Invitrogen) and 120 µg of $Ni^{2+}$-NTA agarose-purified protein in EKMax buffer (Invitrogen). To remove enterokinase from the digest, EK-Away (Invitrogen) was used as recommended by the manufacturer.

### II.5.3.3 $(His)_6$-tagged MetA and MetB

All centrifugations were done in a in a Megafuge 1.0 R (Heraeus) or in a Biofuge fresco (Heraeus). Cells harvested from 150 ml culture volume were resuspended in 150 ml buffer B1, recentrifuged and resuspended in 5 ml buffer B2. Cells were broken as described in II.5.3.1 to obtain 6-8 ml of bacterial lysate, which was cleared by centrifugation (10 min, 6000xg, 4°C). The lysate and 0.15 ml (bed volume) $Ni^{2+}$-NTA-agarose (QIAGEN) equilibrated with buffer B1 were distributed over 2 15-ml polypropylene tubes and incubated horizontally while slowly rocking (15 h, 70 rpm, 4°C) to allow proteins to bind to the matrix. Subsequently, unbound proteins were removed by centrifugation (2 min, 1000xg, 4°C) and removal of the supernatant. The matrix was then washed by agitation in 4 ml buffer B3, recentrifuged, resuspended in 2 ml buffer B3, and the slurry was transferred into a 1-ml polypropylene column (QIAGEN). The settled matrix was washed with 3x 1 ml buffer B4 and 3x 0.3 ml buffer B5. Proteins were then eluted with 3x 0.3 ml buffer BE. The eluate was concentrated as described to obtain 0.4 ml and dialyzed for 15 h at 4°C against buffer BD. Aliquots of dialyzed protein were stored at -20°C for use within 6 weeks. Protein preparations were analysed by SDS-PAGE; a molecular weight of 43 and 41 kDa was assumed for recombinant MetA and MetB, respectively, to approximate the molar concentration of proteins in protein preparations.

| | |
|---|---|
| Buffer B1: | 100 mM Tris/HCl, pH 7.8; 500 mM NaCl; 10 mM imidazole |
| Buffer B2: | Buffer B1 supplemented with 1 mM PMSF, 1 mM benzamidine, 0.5 mM DTT |
| Buffer B3: | 20 mM Tris/HCl, pH 7.0; 500 mM NaCl; 10 mM imidazole |
| Buffer B4: | 20 mM Tris/HCl, pH 7.0; 500 mM NaCl; 20 mM imidazole |
| Buffer B5: | 20 mM Tris/HCl, pH 7.0; 500 mM NaCl; 40 mM imidazole |
| Buffer BE: | 20 mM Tris/HCl, pH 7.0; 500 mM NaCl; 200 mM imidazole |
| Buffer BD: | 20 mM Tris/HCl, pH 7.8, 50 mM NaCl; 0.5 mM EDTA; 1 mM DTT; 50% Glycerol |

## II.6 Electrophoretic mobility shift assay

### II.6.1 Gel mobility shift probes

Arabidopsis mtDNA fragments containing promoters P*atp9*-239 and P*atp9*-295 (see Figure 19) were PCR-amplified from genomic DNA, purified via QIAquick spin columns (QIAGEN), and 5' end-labelled using T4 polynukleotidkinase (PNK, Fermentas). Labelling reactions containing 400 ng PCR product, 10 U T4 PNK, 50 µCi [$\gamma$-$^{32}$P]-ATP (3000 Ci/mmol) (Amersham Biosciences) and 4,8% (w/v) polyethylene glycol 6000 in PNK reaction puffer B (Fermentas GmbH) were incubated for 30 min at 37°C. Reactions were passed over MicroSpin G-50 Columns (Amersham Biosciences) in order to remove unincorporated nucleotides. DNA labelling was monitored using a Bioscan QC 2000 counter.

### II.6.2 DNA binding assay

Binding reactions were set up in a total volume of 30 µl containing 10 mM Tris/HCl (pH 8.0), 10 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 µg/ml BSA /ml, and the indicated amounts of recombinant MetA or MetB protein (50 or 100 ng). Reactions were pre-incubated for 10 min at room temperature, and then supplemented with ~5 ng (~ $10^4$ cpm) of the labelled probe and incubated for another 30 min.

In competition experiments, 5-10 ng of poly(dI-dC) were added after 5 min of pre-incubation. Reaction mixtures were resolved by native PAGE (see II.3.3.4).

## II.7 *In vitro* transcription

### II.7.1 Template construction

mtDNA fragments were PCR-amplified from total Arabidopis DNA with primer pairs listed in Table 6. PCR products were *Sal*I/*Pst*I-digested and ligated into *Sal*I/*Pst*I-cleaved pKL23 (Liere and Maliga, 1999) upstream of terminator sequences to produce plasmid templates for *in vitro* transcription assays (see Figure 23, Figure 25 and Figure 27). Cloned templates were purified from *E. coli* using the QIAGEN Plasmid Midi Kit. Linearized templates were generated by a *Xho*I or *Eco*RI digest and purified via QIAquick spin columns (QIAGEN). *In vitro* transcription templates with their encoded promoters and oligonucleotides employed for plasmid construction are listed in Table 6.

**Table 6: Construction of in vitro transcription templates.** Lowercase nucleotides correspond to non-annealing sequences added in order to introduce restriction sites.

| **Primer pair** | **Primer sequences (5'→3')** | ***In vitro* transcription template** | |
|---|---|---|---|
| | | Name | Promoters |
| pKL23-atp6-1-A-F<br>pKL23-atp6-1-A-R | cagcgagctcCACCGCACGAGTCAGACCT<br>cagcctgcagTCCAGACAGCTTCACTCCGTC | pKL23-*atp6-1*-A | P*atp6-1*-156<br>P*atp6-1*-200 |
| pKL23-atp6-1-B-F<br>pKL23-atp6-1-B-R | cagcgagctcGTTCTGCTTGATTAGGCGAATGC<br>cagcctgcagCGGTTCATCGCCTTACTTATCCA | pKL23-*atp6-1*-B | P*atp6-1*-916/913 |
| pKL23-atp6-1-C-F<br>pKL23-atp6-1-B-R | cagcgagctcCGGGATCAAACTATCAATCTCATA<br>see above | pKL23-*atp6-1*-C | P*atp6-1*-156 |
| pKL23-atp6-2-F<br>pKL23-atp6-2-R | cagcgagctcGGTTCTCCTCTCAGTTCCGTCTA<br>cagcctgcagGTAGCATCCCGCCGATCT | pKL23-*atp6-2* | P*atp6-2*-436<br>P*atp6-2*-507 |
| pKL23-atp8-F<br>pKL23-atp8-R | cagcgagctcCCTGTACATACAAAGATCTAGGCAGC<br>cagcctgcagAACAAAAGCATGGGAGAAAACC | pKL23-*atp8* | P*atp8*-157<br>P*atp8*-228/226 |
| pKL23-atp9-B-F<br>pKL23-atp9-B-R | cagcgagctcTGCGGAAGGAGATTGGAA<br>cagcctgcagGTAGATCATTCGACGTCAGAGGG | pKL23-*atp9*-B | P*atp9*-239<br>P*atp9*-295 |
| pKL23-atp9-F<br>pKL23-atp9-R | cagcgagctcCTTTGGATAATGGTCTAGCCGAGT<br>cagcctgcagTGACAACCTCTAGGGCCAAG | pKL23-*atp9* | P*atp9*-487<br>P*atp9*-652 |
| pKL23-atp9-C-F<br>pKL23-atp9-R | cagcgagctcAGAGAAGGGCAGCATTTATGAGT<br>see above | pKL23-*atp9*-C | P*atp9*-239 |
| pKL23-cox2-F<br>pKL23-cox2-R | cagcgagctcGTTGCCTTGCCTTACCACACC<br>cagcctgcagAGATCACTCTCCTAAAAGCAGCAGTC | pKL23-*cox2* | P*cox2*-210<br>P*cox2*-481 |
| pKL23-rps3-F<br>pKL23-rps3-R | cagcgagctcCAGTCCACCAATAGCGGAAGA<br>cagcctgcagAGATAGAAATGATAGAGGGCCAACC | pKL23-*rps3* | P*rps3*-1053<br>P*rps3*-1133 |
| pKL23-rrn18-F<br>pKL23-rrn18-R | cagcgagctcAGTTGCTTATCCAGGCTTGGTGTT<br>cagcctgcagGCGTACTACTTCCCAACCTTCTGTG | pKL23-*rrn18* | P*rrn18*-69<br>P*rrn18*-156 |
| pKL23-rrn18-C-F<br>pKL23-rrn18-R | cagcgagctcAGAAGGCTGCTTAGAGGAGTGATCT<br>see above | pKL23-*rrn18*-C | P*rrn18*-69 |
| pKL23-rrn26-F<br>pKL23-rrn26-R | cagcgagctcAAAGGCGTTATTGCTGTGCTTCC<br>cagcctgcagCCGCCTCGAATCAAAACGTTC | pKL23-*rrn26* | P*rrn26*-893 |
| pKL23-trnM-F<br>pKL23-trnM-R | cagcgagctcGATTGATTCAATGAAAGTCCC<br>cagcctgcagCCGCTTCTTCTCTTCTACAAG | pKL23-*trnM* | P*trnM*-98 |
| pKL23-trnM-B-F<br>pKL23-trnM-B-R | CCACGGGATTGAGTGAACGAG<br>cagcctgcagGAAGTGAAGCAAGCGAGCCTCT | pKL23-*trnM*-B | P*trnM*-574/573 |

### II.7.2 *In vitro* transcription assay

Standard *in vitro* transcription assays were carried out for 45 min at 30°C and essentially followed the protocol of (Falkenberg, et al., 2002). Reactions contained 6.7 mM Tris/HCl (pH 7.9), 6.7 mM KCl, 6.7 mM $MgCl_2$, 0.67 mM DTT, 0.067% (w/v) BSA, 267 µM each ATP, CTP and GTP, 13 µM unlabelled UTP and 10 µCi of [$\alpha$-$^{32}$P]-UTP (3000 Ci/mmol), 24 U RNase inhibitor and 200 ng of template DNA in a final volume of 15 µl. Reactions were started by adding 400 fmol of recombinant

RpoTm or RpoTmp, and 400 fmol of MetA or MetB where indicated. Reactions were stopped by adding 115 µl RNA extraction buffer (6 M urea, 360 mM NaCl, 20 mM EDTA, 10 mM TRis/HCl pH 8.0, 1% (w/v) SDS) and 20 µl 2.25 M sodium acetate (pH 5.2). Nucleic acids were extracted with phenol/chloroform/isoamyl alcohol (25:24:1), precipitated from the aqueous phase by adding 375 µl ethanol, and washed with 70% (v/v) ethanol. Transcripts were then dissolved in formamide buffer (95% (v/v) formamide; 0,02% (w/v) bromphenol blue; 0,02% (w/v) xylene cyanol) and resolved by denaturing PAGE (II.3.3.3).

### II.7.3 5'-end mapping of *in vitro*-synthesized RNAs

To map 5' ends of *in vitro*-synthesized transcripts, transcription assays were carried out as described in II.7.2, omitting radiolabelled UTP. Following purification, samples were dissolved in ultrapure water and subjected to TAP-treatment and 5'-RACE as indicated in II.3.7; TAP and ligation reactions were downscaled to 1/10 compared the previously described protocol. 5'-RACE performed on non-ligated transcripts served as a control. Reverse primers P2hisa and P3hisa (Table 7) annealing to the *hisa* attenuator sequence in pKL23 derivatives (see Figure 23) were used for cDNA synthesis and PCR (40 cycles), respectively; P1a (see II.3.7) served as forward primer in PCR reactions.

**Table 7: Primers used for 5'-end mapping of in vitro-synthesized RNAs**

| Primer | Primer sequence (5'→3') |
|---|---|
| P2hisa | CACATCGCCTGAAAGACT |
| P3hisa | GGATGATGGTGATGATGGTGG |

## II.8 Green fluorescent protein (GFP) import assay

### II.8.1 GFP targeting constructs

To generate the pMetA-GFP and pMetB-GFP constructs driving the expression of fusion proteins MetA-GFP and MetB-GFP, sequences encoding the 64 N-terminal amino acids of MetA and the 57 N-terminal amino acids of MetB were amplified from reverse-transcribed mRNA-enriched Arabidopsis RNA using primer pairs gfp-metA-P/gfp-metA-M and gfp-metB-P/gfp-metB-2M, respectively (Table 8). PCR products were *Xba*I/*Sal*I-digested and inserted into the *Spe*I/*Sal*I-cleaved vector pOL−GFP S65C (Peeters, et al., 2000). Control constructs encoding mitochondrial CoxIV-GFP and plastidial RecA-GFP (Peeters, et al., 2000) were kindly provided by I. Small (INRA CNRS, Evry, France).

**Table 8: Oligonucleotides used for amplification of MetA and MetB N-termini.** Lowercase nucleotides correspond to non-annealing sequences added in order to introduce restriction sites.

| Primer pair | Primer sequences (5'→3') | Plasmid |
|---|---|---|
| gfp-metA-P | cagctctagaATGATTCTTCGATTGAAAGACCA | pMetA-GFP |
| gfp-metA-M | cagcgtcgacTGCACAGAAACAATCCATCG | |
| gfp-metB-P | cagctctagaATGGCGGGAGGCAAGATC | pMetB-GFP |
| gfp-metB-2M | cagcgtcgacTCACATCGGTACTCTTGATACCAGC | |

### II.8.2 Transient expression in tobacco protoplasts and microscopy

Protoplasts were prepared from leaves of *Nicotiana tabacum* (var. SNN) and transformed with 60 ng of GFP fusion constructs following the protocol of (Morgan and Ow, 1995). Transformed protoplasts were incubated at 20°C in the dark for 16 h prior to microscopy. Epifluorescence microscopy was done using an Axioscope (Zeiss) with GFP- (Zeiss filter set 488013; excitation 470/20, emission 505−530) and FITC- (Zeiss filter set 488009; excitation 450−490, emission LP 520) filter sets.

## II.9 Alignments and phylogeny

Genomic sequences, EST sequences and amino acid sequences were retrieved from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) employing the blastp and tblastn algorithms, and from the *Populus trichocarpa* genome assembly 1.0 (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). Amino acid sequences were derived from genomic and cDNA sequences using the FEX tool at www.softberry.com and the Translate tool at www.expasy.ch. Amino acid sequences were aligned using the Multalin algorithm (Corpet, 1988).

To reconstruct the phylogeny of mitochondrial transcription factors and rRNA dimethylases from an amino acid sequence alignment of these proteins, three different methods were used. For a reconstruction based on Bayesian statistics, the MrBayes program version 3.1 (Ronquist and Huelsenbeck, 2003) was used. The Bayesian inference method employed the JTT amino acid replacement model (Jones, et al., 1992) and a gamma distribution to represent among-site rate heterogeneity (JTT +$\gamma$). A discrete gamma distribution with four categories was assumed to approximate the continuous function. The Metropolis-coupled Markov chain Monte Carlo analysis (MCMC) was performed with 2 million generations and four independent chains. The Markov chain was sampled every 100 generations. Convergence was judged by plots of maximum likelihood (ML) scores and by using the run statistics. The MCMC analysis was assumed to have reached the convergence state if all acceptance rates for the moves in the "cold" chain were in the range 10%–70% and if the acceptance rates for the swaps between chains were also in the range 10%–70%. The first 10000 trees were discarded; the remaining trees were used to construct a consensus tree and to calculate the posterior branch support values. In addition, maximum likelihood and and maximum parsimony analysis were conducted.

## II.10 Material

Ultrapure water was obtained from a USF Purelab Plus system. Chemicals and biochemicals were generally purchased from Roth, ICN Biomedical, Serva, Sigma or Becton-Dickinson; radiochemicals were provided by Amersham Buchler. Deoxyribonucleoside triphosphates and ribonucleoside triphosphates were obtained from Fermentas. Oligonucleotides listed in Tables 1-3 and 5-8 were purchased from Sigma or Eurogentec (annealing temperatures were determined using the oligo calculator at http://www.genscript.com/cgi-bin/tools/primer_calculation). All other materials have been specified in the previous sections.

## II.11 Providers

| | |
|---|---|
| Applied Biosystems | Applied Biosystems, Weiterstadt, Germany |
| Ambion | Ambion, Inc., Austin, USA |
| Amersham | Amersham Buchler GMBH & Co. KG, Braunschweig, Germany |
| Amersham Biosciences | Amersham Biosciences Europe GmbH, Freiburg, Germany |
| Biometra | Biometra GmbH, Göttingen, Germany |
| Bio-Rad | Bio-Rad Laboratories, Richmond, VA, USA |
| Biozym | Biozym Diagnostik GmbH, Hameln, Germany |
| Clontech | Clontech Laboratories, Heidelberg, Germany |
| DuPont | DuPont de Nemours GmbH, Bad Homburg, Germany |
| Epicentre | Epicentre Biotechnologies, Madison, WI, USA |
| Eurogentech | Eurogentech, Seraing, Belgium |
| Fermentas | Fermentas GmbH, St. Leon-Rot, Germany |
| Heraeus | Heraeus, Hanau, Germany |
| Invitrogen | Invitrogen GmbH, Karlsruhe, Germany |
| Macherey-Nagel | Macherey-Nagel, Düren, Germany |
| Millipore | Millipore Corp., Bedford, USA |
| Promega | Promega Corp., Madison, USA |
| QIAGEN | QIAGEN, Hilden, Germany |
| Retsch | Retsch GmbH, Haan, Germany |
| Roche | Roche Molecular Biochemicals, Mannheim, Germany |
| Roth | C. Roth GMBH & Co, Karlsruhe, Germany |
| Serva | Serva Feinbiochemika, Heidelberg, Germany |
| Sigma | Sigma Chemical Company, St. Luis, USA |
| Stratagene | Stratagene, La Jolla, CA, USA |
| USF | USF, Seral Reinstwassersysteme GmbH, Germany |
| Whatman | Whatman Paper, Maidstone, UK |

# III RESULTS

## III.1 Analysis of mitochondrial promoters in *Arabidopsis thaliana*

### III.1.1 Identification of transcription initiation sites by 5'-RACE

To learn about promoter specificities of the mitochondrial transcription machinery in Arabidopsis, mitochondrial transcription initiation sites were experimentally determined using a 5'-RACE technique first described by Bensing et al. (1996) (Figure 7), which since has been applied to define primary transcript 5' termini in different groups of bacteria (Argaman, et al., 2001; Vogel, et al., 2003) and in plastids (Miyagi, et al., 1998). In bacteria as in mitochondria and plastids, primary transcript 5' ends carry triphosphates while processed transcripts have monophosphates at their 5' ends. Only the latter are a substrate to RNA ligase, and are in the experimental procedure selectively ligated to an RNA oligonucleotide, to which a forward primer will anneal in a subsequent 5'-RACE step. Primary 5' termini may be ligated only after removal of a 5' pyrophosphate through tobacco acid pyrophosphatase (TAP). Consequently, 5'-RACE will yield products from TAP-treated RNA for both primary and processed transcripts, whereas without exposure to TAP, products resulting from primary transcript termini will be significantly reduced or absent. Comparison of 5'-RACE products obtained from TAP-treated and untreated RNA (lanes +T and –T in Figure 8 and Figure 11) would thus identify primary transcripts.

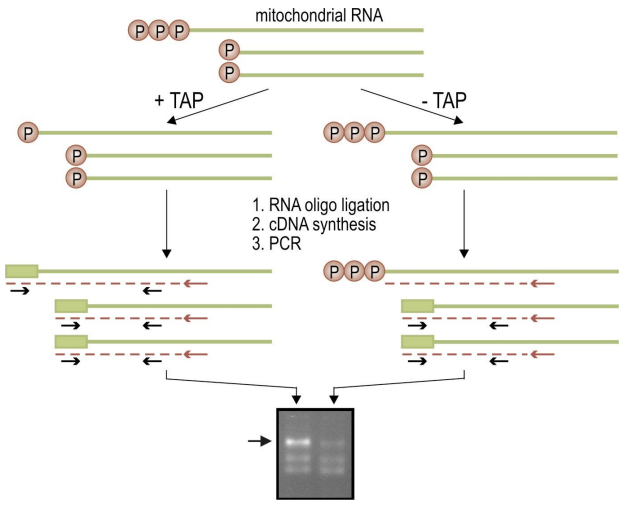



**Figure 7: 5'-RACE technique used to distinguish primary from processed transcript 5' termini**. Transcripts are exposed to TAP to convert 5' triphosphates to monophosphates (left), or not treated with TAP in a control experiment (right). An RNA linker (green box) is then ligated to the 5' monophosphate ends, and cDNA (red dashed lines) is synthesized using a primer complementary to the gene of interest (red arrows). Reverse-transcribed ligation products are amplified using a forward primer annealing to the linker sequence and a gene-specific nested reverse primer (small black arrows). RT-PCR products are analyzed by agarose gel electrophoresis, and products derived from primary transcripts (band indicated by an arrow) are identified by comparing TAP-treated and untreated samples as detailed in the text. After Bensing et al. (1996).

The only transcription start site that has so far been experimentally defined in Arabidopsis mitochondria is located upstream of the *rrn18* gene and coincides with a conserved nonanucleotide sequence motif (Giese, et al., 1996). Including *rrn18* as a control, the genes *rrn18*, *cox2* and *atp9* were first investigated for which promoters have been characterized in several dicots (Binder, et al., 1995; Brown, et al., 1991; Giese, et al., 1996; Lizama, et al., 1994). To look for possible tissue-specific variations in promoter utilization, analysis of transcript 5' ends was performed on RNA isolated from leaves and from flowers of Arabidopsis plants.

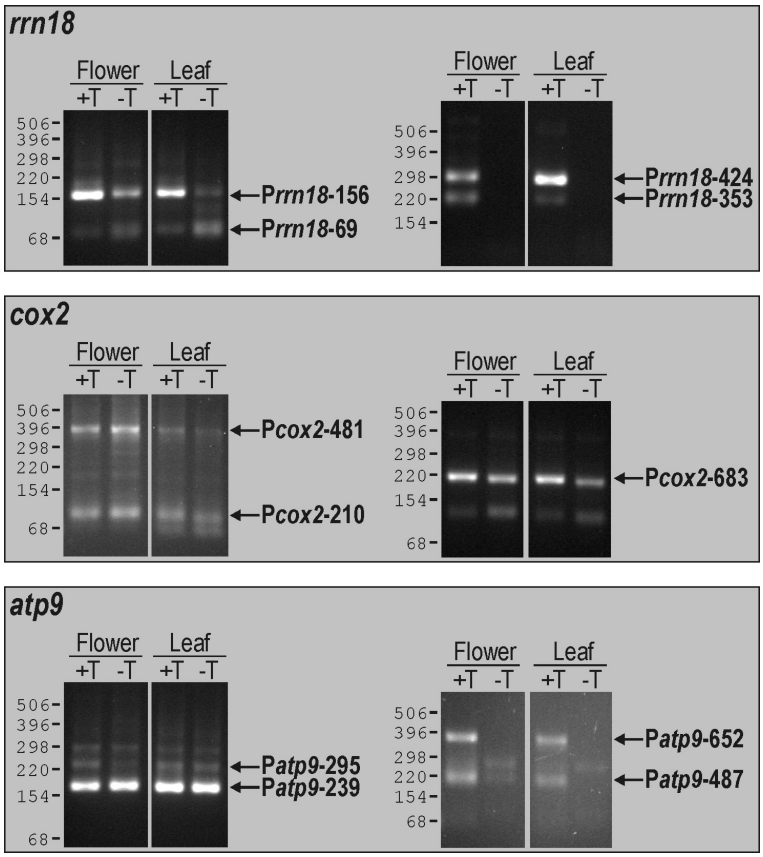



**Figure 8: 5'-RACE analysis of the mitochondrial *rrn18*, *cox2* and *atp9* transcripts.** 5'-RACE products were separated on agarose gels alongside molecular weight markers; sizes are given in base pairs (marker lane not displayed). Products corresponding to primary transcript 5' ends are indicated by arrows and labelled with the name of the corresponding promoter as listed in Table 9. In most cases, TAP-specific products were obtained, which correspond to 5' termini resulting from transcription initiation (lanes designated +T; see also Figure 10 and Table 9). Initiation at P*atp9*-239 and P*atp9*-295 required confirmation through ribonuclease protection analysis of cap-labelled transcripts (compare Figure 10 and Figure 12). Control experiments were done by 5'-RACE from RNA mock-treated in TAP buffer without TAP (lanes –T).

An *rrn18* transcription start site was identified that mapped to position -156 with respect to the mature 18S rRNA 5' end (Figure 8 and Figure 9). The initiating nucleotide was found to be part of the motif CGTATATAA (initiating nucleotide underlined), which has not yet been described as a promoter motif. The previously determined primary end of this transcript at

position -69 (Giese, et al., 1996) appeared to result from processing rather than transcription initiation, as it gave rise to a PCR product that was not enhanced after TAP treatment of transcripts, compared with the control (Figure 8 and Figure 9). In the following, transcriptional starts and their surrounding sequences, which in plant mitochondria encompass the promoter (Caoile and Stern, 1997; Dombrowski, et al., 1999; Rapp, et al., 1993; Rapp and Stern, 1992), will be specified with the letter P ("promoter"), followed by the gene name and position of the initiating nucleotide with respect to the start of the coding sequence or the mature RNA, e.g. P*rrn18*-156.

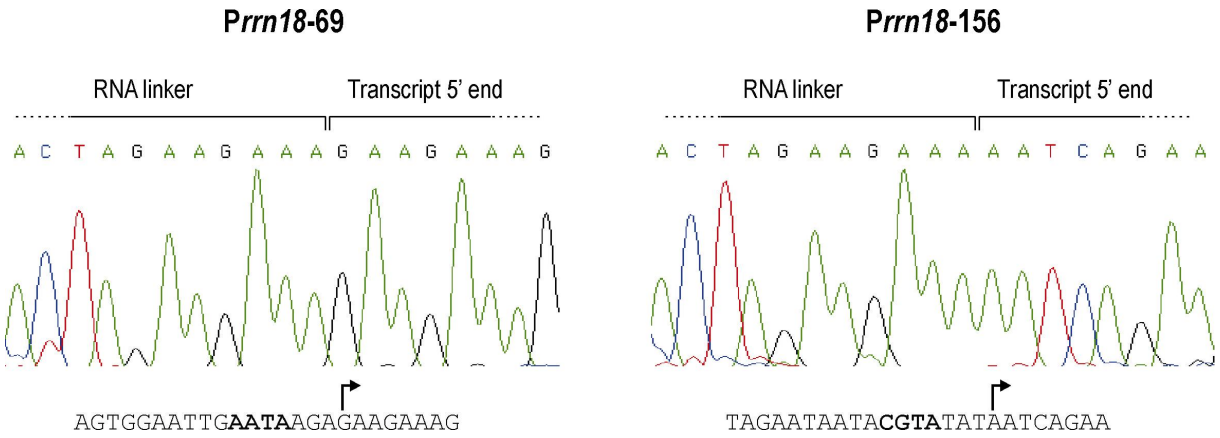



**Figure 9: 5' end identification by sequencing across ligation sites of 5'-RACE products.** Chromatograms display the sequences at ligation sites of typical cloned 5'-RACE products derived from transcripts initiated at P*rrn18*-156 and P*rrn18*-69 (see Figure 8); RNA linker and transcript portions of sequences are indicated. The mtDNA sequences at P*rrn18*-156 and P*rrn18*-69 are displayed below; bent arrows indicate transcription initiation sites.

In the *cox2* upstream region, two transcriptional starts were detected by 5'-RACE. Although TAP-treated and non-treated RNAs lead to similar band patterns (Figure 8), extensive sequencing of cloned PCR products revealed that among products of similar lengths, particular 5' ends were significantly enriched or exclusively present in the TAP-treated sample (Table 9 and Figure 10) and are thus bona fide primary ends. While a nonanucleotide sequence at P*cox2*-210 matched the motif found at P*rrn18*-156 exactly, only limited similarity to any known plant mitochondrial promoter was seen for P*cox2*-481.

5'-RACE analysis of *atp9* transcripts identified one major and one minor 5' end, the latter mapping to position -295 within the motif CGTATATAA and the former mapping to position -239 within the sequence CATAAGAGA which, based on sequence comparisons with the experimentally defined *atp9* promoter in pea mitochondria, had been predicted to function as a promoter upstream of *atp9* and several other genes in Arabidopsis mitochondria (Dombrowski, et al., 1998). However, PCR products resulting from either 5' end were equally

abundant after amplification from TAP-treated and non-treated RNA (Figure 8), and transcripts were found to start with the nucleotides underlined in Figure 10, regardless of the application of TAP (Figure 10). Thus, both 5' ends were carrying 5' monophosphates and therefore resulting from processing events, despite the perfect nonanucleotide motifs. Alternatively, mixed populations of primary and processed transcript 5' ends might have been present, starting with identical nucleotides but carrying either tri- or monophosphates. In order to unambiguously determine whether transcription initiated at positions -239 and -295, the species of *atp9* transcript ends mapping to these positions were tested for the presence of *in vitro*-cappable 5' termini (see below).

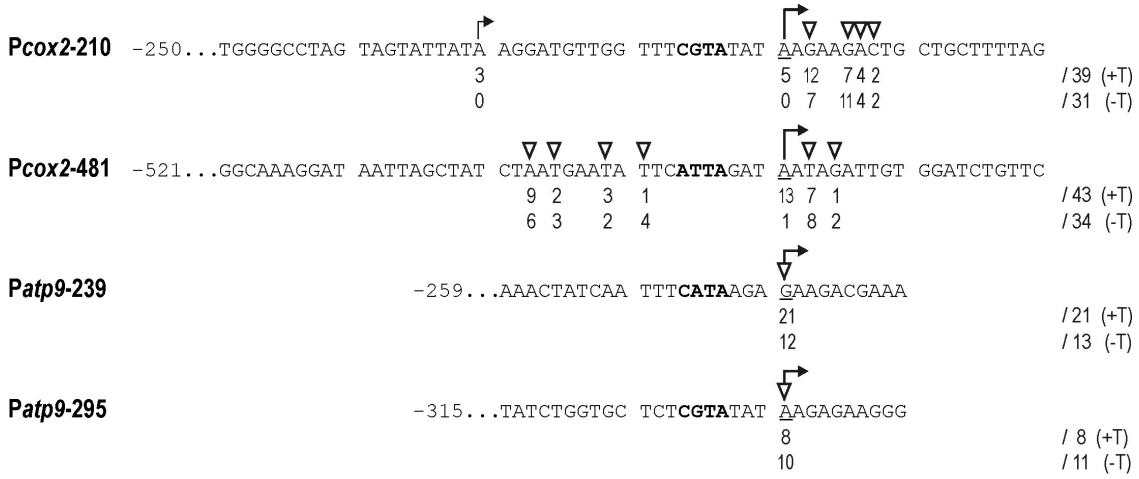



**Figure 10: Transcript 5' termini detected through cloning and sequencing of 5'-RACE products designated P*cox2*-210, P*cox2*-481, P*atp9*-239 and P*atp9*-295 in Figure 8.** Parts of the *cox2* and *atp9* upstream sequences that surround the four transcription initiation sites are shown; numbers preceding the sequences are the positions of the first displayed nucleotide with respect to the translational start. Numbers written below nucleotide positions indicate frequencies of clones that were found to correspond to transcript 5' ends mapping to the respective positions (row +T, clone numbers for products of 5'-RACE following TAP treatment; row –T, clone numbers determined without TAP treatment). Only 5' ends detected more than twice are marked. Numbers behind slashes indicate the numbers of clones that were sequenced in total for each promoter. Nucleotides corresponding to 5' ends that most likely result from processing events are indicated by open triangles. Transcription initiation sites, which gave rise to TAP-specific 5'-RACE products, are indicated by bent arrows. Processing sites and initiation sites were appointed as detailed in the text. Upstream of P*cox2*-210, a small bent arrow marks position -231, which might be a transcription initiation site but did not yield a distinct band in 5'-RACE experiments.

With the aim of identifying additional promoters of the genes *rrn18*, *cox2* and *atp9*, their 5' regions were analyzed by 5'-RACE through repeatedly placing reverse primers upstream of identified transcriptional starts, until no further transcript ends could be detected. All three genes were found to possess additional upstream promoters (right panels in Figure 8), of which none matched any known plant mitochondrial promoter sequence (Table 9).

The Arabidopsis mitochondrial genome was screened for additional occurrences of sequence motifs coinciding with experimentally defined transcriptional starts. Of the genes

displaying a promoter motif in their upstream regions, *rrn26*, *atp1*, *atp6-1*, *atp6-2* and *atp8* were selected for an experimental verification of their predicted transcriptional starts. Notably, the *atp1* and *atp6-1* 5' regions like *atp9* displayed the motif CGTATATAA approximately 50 base pairs upstream of the hypothetical CATAAGAGA promoter sequence. The analysis moreover included *tRNA-fMet* with the predicted promoter motif CGTAAGAGA (Dombrowski, et al., 1998), which had been found to be an element of the *atp9* promoter in pea and soybean (Binder, et al., 1995; Brown, et al., 1991). Of those genes not possessing any conserved promoter motif upstream of their coding sequence, *rps3* and *cox1* were selected for transcript 5' end mapping.

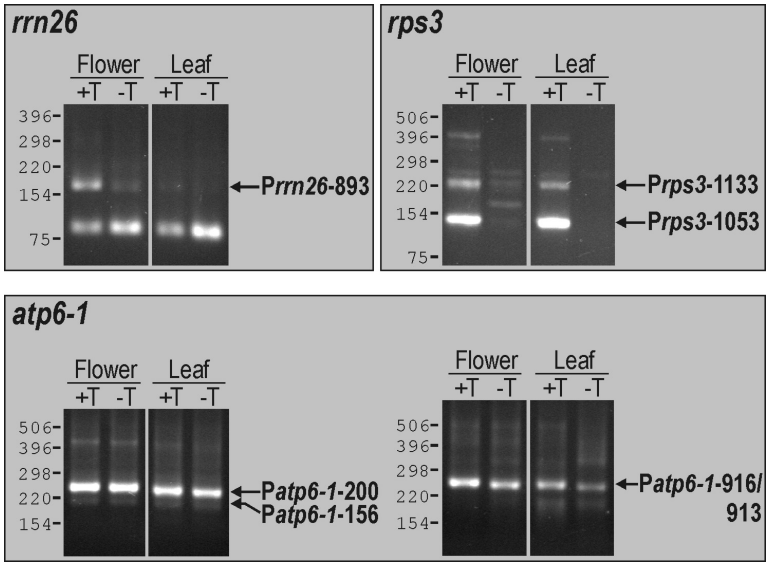



**Figure 11: 5'-RACE analysis of the mitochondrial *rrn18*, *rps3* and *atp6-1* transcripts.** Amplified products were separated on agarose gels alongside molecular weight markers; sizes are given in nucleotides (marker lane not displayed). TAP-specific products (lane +T) that correspond to primary transcript 5' ends are indicated by arrows and labelled with the name of the respective promoter as listed in Table 9. Initiation at P*atp6-1*-156 and P*atp6-1*-200 required confirmation through ribonuclease protection analysis of cap-labelled transcripts (compare Figure 12). Control experiments were done by 5'-RACE from RNA mock-treated in TAP buffer without TAP (lanes –T).

Table 9 provides a summary of determined transcription start sites and their surrounding sequences. Besides 5' ends that were unambiguously identified as primary ends in the 5'-RACE, such as those mapping to P*rrn26*-893 and P*rps3*-1053, various 5' termini were detected for which 5'-RACE products were not enhanced following TAP treatment of RNAs but which, as described above for P*cox2*-210, nevertheless coincided with genomic sequences exhibiting strong similarity to bona fide promoters. For these transcripts the pools of 5' termini cloned from +TAP and from –TAP samples were again compared (fourth and fifth column in Table 9). Mostly, the –TAP pool contained slightly shorter transcripts than the

+TAP pool (for examples, compare sizes of +TAP and –TAP 5'-RACE signals obtained for P*atp6-1*-916/913 and P*cox2*-683 in Figure 11 and Figure 8, respectively), and particularly was deprived of longer transcript species that at their 5' extremities carried A or G nucleotides (see Table 9). These transcripts specific to the +TAP pool are most likely resulting from transcription initiation. Within three promoter regions located upstream of the *tRNA-fMet*, *atp6-1* and *atp8* genes, transcription was found to initiate at two different nucleotide positions. From 5'-RACE results it is likely that multiple initiations also occur around P*rps3*-1133 (data not shown). Multiple promoters were detected for all investigated genes except *rrn26, cox1* and *orf291*. Due to partly identical upstream and coding sequences of *cox2* and *orf291*, the transcriptional start site preceding *orf291* was fortuitously found using primers annealing to the *cox2* upstream region.

To analyze possible differences in promoter utilization between Arabidopsis leaves and flowers, TAP-specific 5'-RACE signals (lanes +T in Figure 8 and Figure 11) that had been obtained from leaf and from flower RNA for a distinct gene were compared. No primary transcript 5' end was detected that was exclusively present in leaves or in flowers, indicating that transcription is initiated at identical sites in both tissues. An occasional enhancement of 5'-RACE signals from flower RNA can be attributed to the level of mitochondrial activity being generally higher in flowers than in green tissues (Huang, et al., 1994; Smart, et al., 1994).

### III.1.2 Identification of transcription initiation sites by *in vitro* capping

As already observed for *atp9*, 5'-RACE analyses of those *atp1*, *atp6-1*, *atp6-2* and *atp8* transcript 5' termini mapping to the motifs CATAAGAGA and CGTATATAA did not support transcription initiation at these sequences (Table 9 and Figure 11). The perfect nonanucleotide motifs found at these sites prompted the examination of the corresponding 5' ends by an independent technique. As a method specifically detecting primary 5' ends, ribonuclease protection of *in vitro*-capped transcripts was employed to analyze the respective 5' termini of the *atp9* and *atp6-1* mRNAs. This method takes advantage of organellar transcripts being, unlike nuclear mRNAs, not capped at their 5' ends *in vivo*. Mitochondrial primary transcripts, which carry 5' triphosphates, are thus representing guanylyltransferase (capping enzyme) substrates and can be 5' cap-labelled with the GMP moiety of [$^{32}$P]–α-GTP *in vitro*. Total Arabidopsis RNA was capped and then subjected to ribonuclease protection using RNA probes complementary to the genomic regions containing putative promoters. The *rrn26* primary transcript was included as a positive control in the capping study, since its 5'

end had been established by 5'-RACE to map to a promoter that is identical to the sequence surrounding the predicted transcriptional start P*atp1*-1947, and moreover is highly similar to the hypothetical promoters P*atp6-1*-156, P*atp6-2*-148, P*atp8*-157 and P*atp9*-239. Additionally, the *rrn18* transcript 5' ends coinciding with positions -69 and -156 were tested for their ability to be capped *in vitro*.

**Table 9: Transcription initiation sites detected by 5'-RACE and *in vitro*-capping.** Initiating nucleotides are underlined; repeatedly observed promoter cores are written bold and the frequent TATATA(A) motif is highlighted. The number of clones that were sequenced for each promoter is given together with the frequency of the respective primary transcript 5' end as determined from TAP-treated flower RNA, and for selected promoters from flower RNA not exposed to TAP.

| Gene | Promoter | Sequence | No. of clones (+TAP) | No. of clones (-TAP) | *In vitro*-cappable |
|---|---|---|---|---|---|
| *rrn18* | P*rrn18*-156 | TAGAATAATA**CGTA**TATAATCAGAA | 20/23 | 4/7 | + |
| *orf291* | P*orf291*-307 | TGGAATAATA**CGTA**TATAATCAGAT | 7/10 | n.d. | n.d. |
| *atp6-1* | P*atp6-1*-200 | GCCAATAATA**CGTA**TATAAGAAGAG | 3/14 | n.d. | + |
| *atp9* | P*atp9*-295 | CTGGTGCTCT**CGTA**TATAAGAGAAG | 8/8 | 10/11 | + |
| *atp1* | P*atp1*-1947 | CTGGTGGTAT**CGTA**TATAAGAGAGA | 8/13 | 10/15 | + |
| *cox2* | P*cox2*-210 | ATGTTGGTTT**CGTA**TATAAGAAGAC | 5/39 | 0/31 | + |
| *tRNA-fMet* | P*trnM*-98 [b] | TTTGAAATAT**CGTA**AGAGAAGAAGG | 12/12 | n.d. | n.d. |
| *rrn26* | P*rrn26*-893 [b] | CTATCAATTT**CATA**AGAGAAGAAAG | 12/13 | 0/23 | + |
| *atp1* | P*atp1*-1898 [b] | CTATCAATTT**CATA**AGAGAAGAAAG | 13/13 | n.d. | + |
| *atp9* | P*atp9*-239 [b] | CTATCAATTT**CATA**AGAGAAGACGA | 21/21 | 12/13 | + |
| *atp6-1* | P*atp6-1*-156 [b] | CTATCAATCT**CATA**AGAGAAGAAAT | 5/14 | n.d. | + |
| *atp6-2* | P*atp6-2*-148 [b] | CTATCAATCT**CATA**AGAGAAGAAAT | 7/13 | n.d. | + |
| *atp8* | P*atp8*-157 [b] | CTATCAATCT**CATA**AGAGAAGAAAT | 14/22 | n.d. | n.d. |
| *rrn18* | P*rrn18*-69 [a] | AGTGGAATTG**AATA**AGAGAAGAAAG | n.d. | 6/8 | + |
| *atp8* | P*atp8*-999 | ATAAAATTA**AATA**AAGAGCAAAAAT | 9/12 | n.d. | n.d. |
| *atp8* | P*atp8*-228/226[c] | CATACCATAA**CATA**TATAGAATCGA | 1/28, 6/28 | 0/14, 0/14 | n.d. |
| *rrn18* | P*rrn18*-353 | TACTTTTCCATCTATATAAAATGAA | 10/12 | n.d. | n.d. |
| *atp6-1* | P*atp6-1*-916/913 [c] | AGCCCTTTAT**ATTA**TATAATAAAGC | 2/23, 11/23 | 0/23, 1/23 | n.d. |
| *cox1* | P*cox1*-355 | AATTTATTCA**ATTA**TATAATAATAA | 18/23 | 19/30 | n.d. |
| *cox2* | P*cox2*-481 | ATGAATATTC**ATTA**GATAATAGATT | 13/43 | 1/34 | n.d. |
| *rps3* | P*rps3*-1133 | TAGAAAAAATT**ATTA**GTAATACGTA | 6/26 | 0/15 | n.d. |
| *rrn18* | P*rrn18*-424 | TCAAATCCTC**GGTA**TATAAAGAGAA | 9/10 | n.d. | n.d. |
| *cox2* | P*cox2*-683 | GACACGTAA**GGTA**AAATAAGAATCT | 6/12 | 0/8 | n.d. |
| *rps3* | P*rps3*-1053 | TTTTTTATTT**GGTA**GGTAACATCGC | 12/14 | n.d. | n.d. |
| *atp9* | P*atp9*-487 | ATGTCTTATT**GGTA**TGTGATACAAG | 13/14 | n.d. | n.d. |
| *atp9* | P*atp9*-652 | AGAAGATTGA**AGTA**AGGAGCAGGTT | 7/16 | 0/29 | n.d. |
| *atp6-2* | P*atp6-2*-436 | TCTTGAATTA**AGTA**TATAGAAAAGA | 5/20 | n.d. | n.d. |
| *atp6-2* | P*atp6-2*-507 | GATAAATTA**AGTA**TAGTAATAAGAA | 9/12 | n.d. | n.d. |
| *atp8* | P*atp8*-710 | ATCGGAGCTGCCAATAAGCTAATCC | 4/12 | 0/13 | n.d. |
| *tRNA-fMet* | P*trnM*-574/573 [c] | CTAATTTATATAAAAAAGACCGGGA | 9/18, 9/18 | n.d. | n.d. |

[a] Consistent with primer extension results in Giese et al. (Giese, et al., 1996).
[b] Consistent with previous predictions of Arabidopsis mitochondrial promoters (Dombrowski, et al., 1998).
[c] Transcription initiation was found to occur at two different nucleotides in one promoter region; frequencies of transcript 5' termini are given first for the upstream nucleotide.
n.d., not determined.

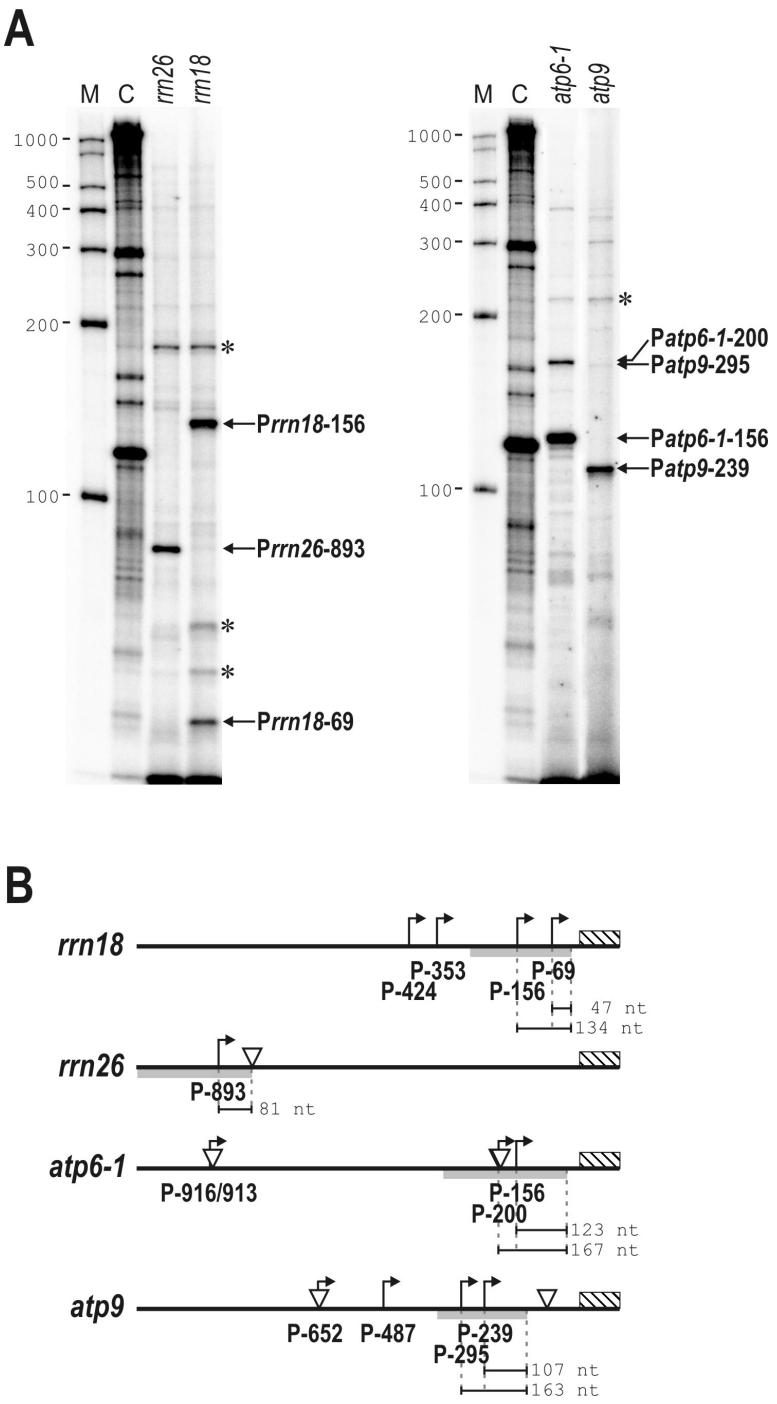



**Figure 12: Detection of selected *rrn18, rrn26, atp6-1* and *atp9* primary transcript 5' ends by ribonuclease protection of cap-labelled RNA.** (**A**) Protected RNA fragments were separated in polyacrylamide gels alongside a molecular weight marker (lane M); sizes are given in nucleotides. Lane C shows total capped RNA prior to ribonuclease protection. Lanes designated *rrn18, rrn26, atp6-1* and *atp9* show ribonuclease protection results obtained with riboprobes complementary to the *rrn18, rrn26, atp6-1* and *atp9* upstream regions as detailed in Figure 12B. Specific protected fragments, which correspond to primary transcripts, are indicated by arrows and labelled with the respective promoter name. Asterisks mark fairly strong signals that were considered non-specific, as they were seen with different riboprobes. (**B**) Diagram of the *rrn18, rrn26, atp6-1* and *atp9* 5'-untranslated regions. Promoters identified in 5'-RACE and capping analyses are indicated by bent arrows; open triangles mark processing sites identified by 5'-RACE. The beginning of the mature rRNA or protein-coding sequence is indicated by a hatched bar. Grey bars are drawn below sequences complementary to the riboprobes that were annealed to cap-labelled primary transcripts in ribonuclease protection assays. The sizes of expected protected RNA fragments are given in nucleotides.

Figure 12A shows the protected cap-labelled RNAs corresponding to the transcription initiation sites P*rrn26*-893, P*rrn18*-69 and P*rrn18*-156, and to the tandem promoters P*atp6-1*-156 and P*atp6-1*-200, and P*atp9*-239 and P*atp9*-295. The sizes of protected RNA fragments are in accordance with the expected lengths of transcript 5' segments annealing to RNA probes, as illustrated in Figure 12B. Most notably, P*rrn18*-69 as well as P*atp6-1*-156, P*atp6-1*-200, P*atp9*-239 and P*atp9*-295, which through 5'-RACE could not be confirmed as transcription initiation sites, were found to coincide with *in vitro*-cappable and thus primary RNA 5' termini. Capping of the *atp9* mRNA mapping to position -295 only yielded a very faint signal, which may be either because of rare utilization of P*atp9* -295 as a promoter, or due to rapid *in vivo* processing of this primary message. *In vitro* capping moreover verified transcription initiation at P*atp1*-1898, P*atp1*-1947 and P*atp6-2*-148 (data not shown). The analyses of *in vitro*-cappable *atp1*, *atp6-1*, *atp6-2* and *atp9* mRNAs allow to infer that the transcript 5' terminus coinciding with P*atp8*-157 is also derived from transcription initiation.

### III.1.3 Mitochondrial promoter architecture in Arabidopsis

Table 9 aligns Arabidopsis mitochondrial promoter sequences with respect to experimentally defined transcription start sites and places promoters with similar core sequences in adjacent rows. At positions -7 to -4 with respect to the transcriptional start, the majority of promoters display the previously described core element CRTA (R = A or G) (Fey and Marechal-Drouard, 1999), which here is almost always seen as part of the nonanucleotide motifs CGTATATAA or CATAAGAGA, or the sequences ATTA, AGTA, GGTA or AATA. Only in a few promoters is the distance between core element and start site altered by one base pair.

All primary transcripts characterized in this study originate from transcription initiation at either an A or a G nucleotide (21 and 12 out of 33 start sites, respectively). When comparing nucleotide frequencies within promoters, it appears that A and G start sites favour distinct nucleotides at adjoining positions. For example, while a G as initiating nucleotide is nearly always preceded by an A, initiation at an A essentially requires a T at position -1. Due to these constraints on nucleotide frequencies particularly at positions around transcription initiation sites, promoter sequences were realigned in two subsets. The two alignments of promoters driving transcription from an A or a G are illustrated in Figure 13 as sequence logos. Among promoters directing initiation at an A, the sequence element TATATA(A) is fairly frequent. Promoters having a G nucleotide at position +1 mostly conform to the

consensus CRTAAGAGA that has been suggested previously for dicot mitochondrial promoters (Binder, et al., 1996).

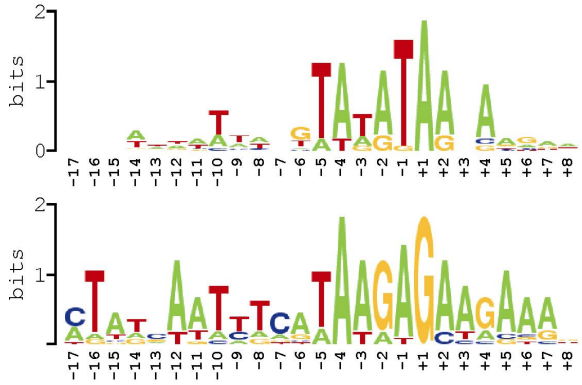



**Figure 13: Summary of nucleotide sequences around experimentally defined transcription initiation sites in Arabidopsis mitochondria, as displayed in** Table 9**.** Two sequence logos are shown that were generated using WebLogo [(http://weblogo.berkeley.edu/logo.cgi, (Crooks, et al., 2004; Schneider and Stephens, 1990)] from an alignment of 20 promoter sequences activating transcription initiation at an adenine nucleotide (upper sequence logo) and from an alignment of 11 sequences supporting initiation at a guanine nucleotide (lower sequence logo). Position +1 corresponds to the transcriptional start.

### III.1.4 Promoters directing transcription of non-coding sequences

To gain an approximate idea of genome-wide promoter distribution in Arabidopsis mitochondria, the mtDNA of this plant was computationally screened for additional occurrences of sequences that are identical to mitochondrial promoter elements determined in the present study. Sequence stretches of various promoters that extended from the initiation site to the core tetranucleotide were used as query (compare Table 9). Potential promoters were detected not only upstream of annotated ORFs but also at sites preceding non-coding sequences of several kilobases and on complementary strands of identified genes. To test whether promoters are active in the Arabidopsis mitochondrial genome that direct the synthesis of presumed non-coding or antisense transcripts, selected motifs emerging from the *in silico* search were tested for promoter function by 5'-RACE as described in III.1.1.

Figure 14 summarizes the mapping of 5' termini of antisense transcripts made to different *nad* genes; sequences at transcription initiation sites are given in Table 10. Intron 3 of *nad4* and intron 4 of *nad5* harbour sequences reminiscent of the P*rps3*-1133 promoter on their antisense strands. A *nad5* antisense transcripts initiated at this motif (designated P*nad5*-AS) was easily amplified from Arabidopsis RNA, whereas the *nad4* antisense transcript mapping to a P*rps3*-1133-like promoter (designated P2*nad4*-AS) was weak. A far more abundant 5'-RACE product was derived from a *nad4* antisense transcript mapping to a P*rps3*-1053-like sequence on the complementary strand of intron 3 (P1*nad4*-AS). A prominent signal was also

obtained for a *nad1* antisense transcript synthesized from a P*cox1*-355-like region (P1*nad1*-AS). An additional *nad1* antisense transcript was detected that is initiated from the complementary strand of intron 4 at a sequence comprising a CRTA element (P2*nad1*-AS). The promoter motif CGTATATAA (compare Table 9) is present in antisense orientation within exon 5 of the *nad2* gene and indeed was found to promote transcription (P*nad2*-AS). A *nad7* antisense RNA mapping to a region on the intron-2 complementary strand that contained typical promoter elements (P2*nad7*-AS) gave rise to an abundant 5'-RACE product. Thus, several active promoters of the Arabidopsis mtDNA drive the synthesis of antisense transcripts to known genes.

**Table 10: Transcription initiation sites associated with antisense and non-coding transcripts.** Initiating nucleotides are underlined; typical promoter elements are written bold. The number of clones that were sequenced for each promoter is given together with the frequency of the respective primary transcript 5' end as determined from TAP-treated RNA.

| Promoter | Sequence | No. of clones (+TAP) |
|---|---|---|
| P1*nad1*-AS | GAGAAATACCTT**ATTA**T<u>A</u>TATAT<u>A</u>T<u>A</u>A | 3,3,4/12 [a] |
| P2*nad1*-AS | CAACTAATCT**CATA**AGT<u>A</u>AACGCCT | 3/4 |
| P*nad2*-AS | TTTCACTAAG**CGTA**TAT<u>A</u>ATAAAAT | 9/12 |
| P1*nad4*-AS | TATCATGGTGAA**GGTA**A<u>**GGTA**</u>ACGC | 4,4/11 [a] |
| P2*nad4*-AS | GCCTTT**ATTAGTA**A**AGTA**<u>A</u>AAGCTTT | 3/6 |
| P*nad5*-AS | TTCTCTATTAT**ATTAGTA**<u>A</u>AAGGGAA | 10/11 |
| P*nad7*-AS | A**CGTA**AGAACT**AGTA**TT<u>G</u><u>A</u>AAGCTA | 4,8/12 [a] |
| $P_{38K}$-nc | TCGATAATAT**CGTA**AGA<u>G</u>AAGAAAA | 11/11 |
| $P_{203K}$-nc | CCATCTATTT**CATA**AGA<u>G</u>AATAAAA | 11/12 |

[a] Transcription initiation was found to occur at different nucleotides in one promoter region; frequencies of transcript 5' termini are given first for the upstream nucleotide.

Of the promoter-like sequences not preceding identified or hypothetical ORFs, a P*trnM*-98-like motif ($P_{38K}$-nc) which would drive initiation at nucleotide 38411 from the reverse strand and a P*rrn26*-893-like motif ($P_{203K}$-nc) around nucleotide 203267 on the direct strand of the Arabidopsis mtDNA were tested by 5'-RACE for promoter function (Figure 15 and Table 10). Transcripts mapping to the predicted transcriptional starts were amplified for both promoters, indicating that regions of the Arabidopsis mtDNA that lack known genes are indeed expressed from individual promoters.

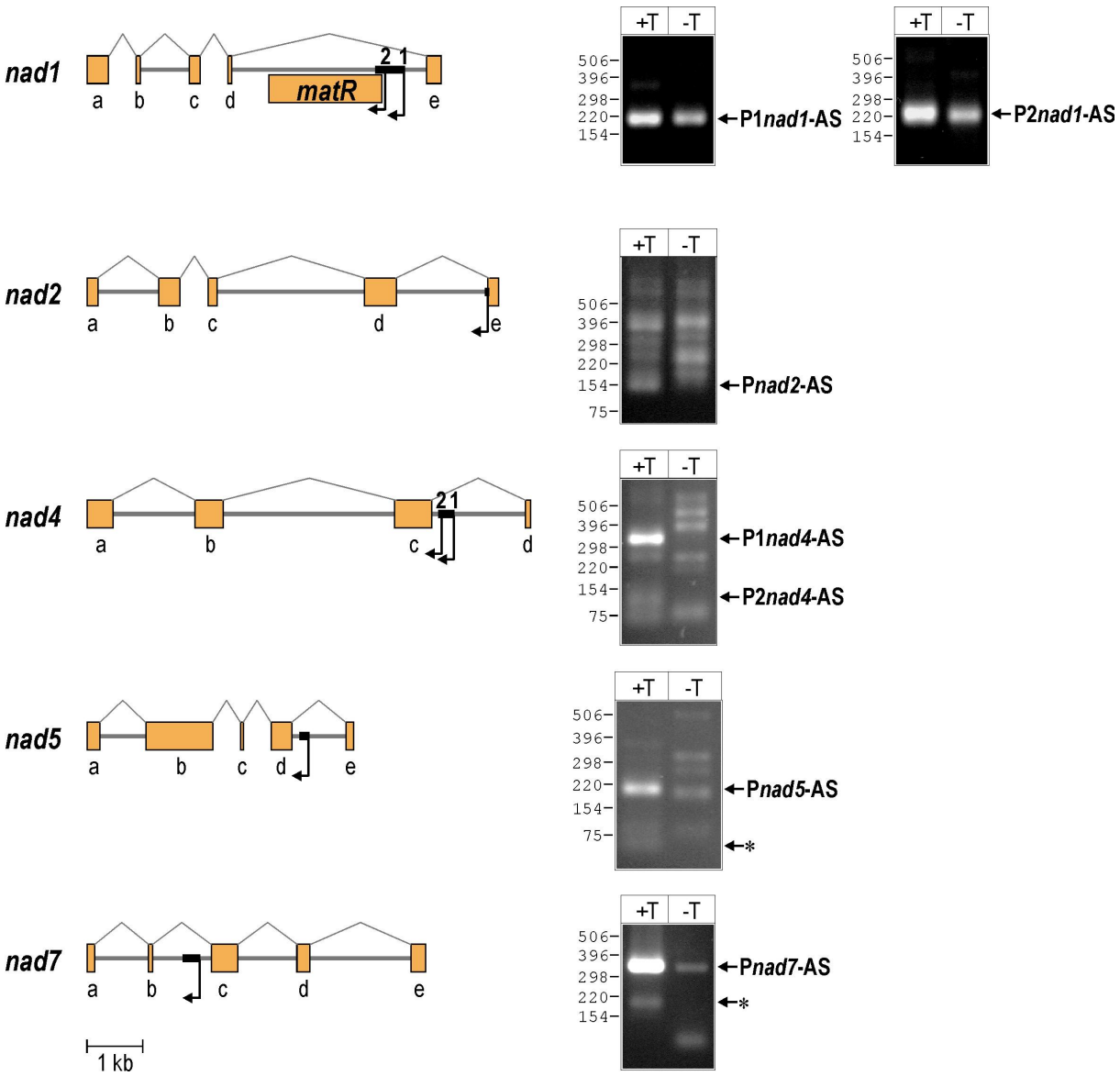



**Figure 14: Synthesis of antisense transcripts to mitochondrial *nad* genes.** Agarose gel analyses of 5'-RACE products derived from *nad* antisense transcripts (right; size marker and lane designation as in Figure 8) are displayed together with diagrams illustrating *nad* gene structures and positions of promoters driving antisense RNA synthesis (left). 5'-RACE signals corresponding to promoters are indicated by arrows beside gels and are specified according to Table 10; asterisks mark RNAs identified by sequencing to represent non-specific products. In *nad* gene diagrams, orange boxes and horizontal grey lines mark exons [exon assignment as in (Unseld, et al., 1997)] and *cis*-spliced introns respectively; the absence of horizontal grey lines between exons indicates trans-splicing (compare Figure 2 for *nad* exon distribution on the Arabidopsis mtDNA). Bent arrows symbolize promoters; gene fragments corresponding to amplified antisense transcripts are represented by bold black lines.

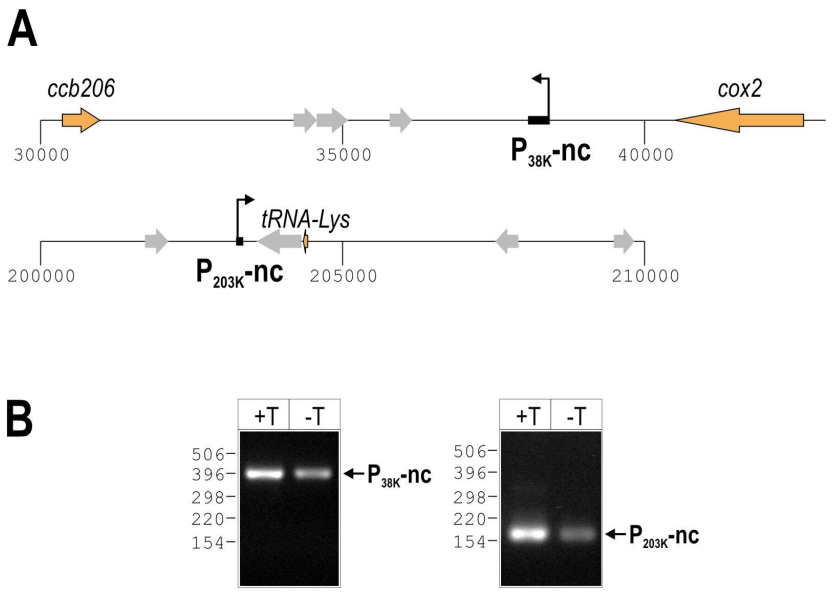



**Figure 15: Synthesis of transcripts from intergenic regions of the Arabidopsis mtDNA.** (**A**) Location of promoters $P_{38K}$-nc and $P_{203K}$-nc driving transcription of non-coding sequences; mtDNA coordinates (in nt) are indicated. Orange arrows denote functional genes and are labelled accordingly; grey arrows correspond to hypothetical ORFs. Bent arrows symbolize promoters; gene fragments corresponding to amplified transcripts are represented by bold black lines. (**B**) Agarose gel analysis of 5'-RACE products derived from transcripts initiated at $P_{38K}$-nc and $P_{203K}$-nc (size marker and lane designation as in Figure 8). Signals corresponding to promoters are indicated by arrows beside gels and are specified according to Table 10.

5'-RACE products displayed in Figure 14 and Figure 15 were obtained using RNA isolated from Arabidopsis flowers; essentially the same signals were obtained for leaf transcripts (data not shown).

## III.2 Characterization of a mitochondrial mtTFB-like protein in Arabidopsis

### III.2.1 Identification of mtTFB-like sequences in the Arabidopsis genome

The present study aims at dissecting the roles of the two phage-type RNA polymerases present in Arabidopsis mitochondria by characterizing the transcriptional performances of RpoTm and RpoTmp *in vitro*. Specific transcription initiation at mitochondrial promoters may require complementing recombinant RpoT enzymes with as yet unidentified auxiliary factors (see I.3.3). Therefore, the Arabidopsis genome was screened for sequences encoding candidate cofactors of phage-type RNA polymerases.

BLAST searches using the amino acid sequence not of mtTFB from *S. cerevisiae* but of a putative *Schizosaccharomyces pombe* mtTFB homologue as query have previously lead to identifying nuclear genes encoding the transcriptional cofactors mtTFB1 and mtTFB2 of human mitochondria (Falkenberg, et al., 2002; McCulloch, et al., 2002). Hence, the Arabidopsis genome and ESTs were queried with the putative *S. pombe* mtTFB amino acid sequence (McCulloch, et al., 2002) and the human mtTFB1 and mtTFB2 sequences (Falkenberg, et al., 2002) using the blastp and tblastn algorithms available at the National Centre for Biotechnology Information. Three mtTFB-like dimethyladenosine transferases are predicted to be encoded by the loci At5g66360, At2g47420 and At1g01860 (see I.3.3.1 for details on the structural similarity of yeast and animal mitochondrial transcription factors to rRNA dimethyladenosine transferases), of which the latter corresponds to the previously characterized *PFC1* gene coding for a plastidial 16S rRNA dimethylase (Tokuhisa, et al., 1998). BLAST searches of the Arabidopsis genome using any of the three sequences as query did not deliver additional hits. No sequences encoding Arabidopsis mtTFA homologues were identified by screening the database for putative mitochondrial HMG box proteins.

The methyltransferase-like genes at loci At5g66360 and At2g47420 were tentatively designated *MetA* and *MetB* (methyltransferase-like), respectively. While *MetB* is predicted to encode a 353-amino acid polypeptide, available EST data support alternative splicing of the hypothetical mRNA deriving from *MetA*, which would give rise to two different polypeptides of 352 and 380 amino acids (GenPept accession numbers NP_201437 and NP_975003). However, PCR amplification of the *MetA* and *MetB* coding sequences from cDNAs yielded only the longer of the two predicted products for *MetA* and a fragment of the expected length for *MetB* (data not shown). Protein sequence comparisons revealed that the presumed optional *MetA* intron codes for amino acid sequence motifs that are conserved among rRNA dimethylases-like proteins (amino acids 214-241 of the derived MetA polypeptide, see alignment in Annex A). Hence, further sequence analyses were based on the longer deduced

MetA polypeptide, and the amplified *MetA* and *MetB* cDNAs were used to construct plasmids for MetA and MetB expression in *E. coli* (see III.2.4).

Both MetA and MetB display ~30% and ~27% amino acid sequence similarity to mtTFB from *S. pombe* and *S. cerevisiae*, respectively; ~15% and ~12 % of positions are identical (sequences exclusive of the predicted transit peptides were compared). These values approximately correspond to those obtained from comparisons of human to fungal mtTFB sequences. Similarities of MetA and MetB to each h-mtTFB1 and h-mtTFB2 are ~37%; identities to the human sequences are ~20% for both MetA and MetB.

**Table 11: Predicted properties of Arabidopsis MetA and MetB.**

| | **MetA** | **MetB** |
|---|---|---|
| **Locus tag** | At5g66360 | At2g47420 |
| **Subcellular localization** | | |
| TargetP 1.1[a] | mitochondrial | not mitochondrial or plastidial |
| Mitoprot[b] | mitochondrial (P=0.59)[b] | non-mitochondrial (P=0.25)[b] |
| Predodar 0.5[c] | mitochondrial | mitochondrial |
| Predodar 1.03[d] | possibly mitochondrial | not mitochondrial or plastidial |
| PsortII[e] | mitochondrial | nuclear |
| iPsort[f] | mitochondrial | not mitochondrial or plastidial |
| **N-terminal transit peptide** | | |
| TargetP1.1[a] | 26 aa | - |
| Mitoprot[b] | 27 aa | 17 aa |
| PsortII[e] | 27 aa | - |
| **Mature protein**[g] | | |
| Length | 354 aa | 335 aa |
| Molecular weight | 40 kDa | 38 kDa |
| pI | 8.6 | 8.5 |

[a] http://www.cbs.dtu.dk/services/TargetP (Emanuelsson, et al., 2000; Nielsen, et al., 1997)
[b] http://www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofilter (Claros and Vincens, 1996) (P, probability of mitochondrial targeting)
[c] http://www.inra.fr/Internet/Produits/Predotar/
[d] http://genoplante-info.infobiogen.fr/predotar/predotar.html (Small, et al., 2004)
[e] http://psort.ims.u-tokyo.ac.jp/form.html
[f] http://hc.ims.u-tokyo.ac.jp/iPSORT/ (Bannai, et al., 2002)
[g] Physicochemical properties were derived using ProtParam (http://expasy.cbr.nrc.ca/tools/protparam.html) from MetA and MetB amino acid sequences lacking the predicted transit peptides.

### III.2.2 Mitochondrial localization of the mtTFB-like protein MetA

Several computer algorithms unambiguously predicted MetA to possess an N-terminal transit peptide mediating the import of the protein into mitochondria (Table 11). In contrast, mitochondrial localization of MetB was supported only by the older, less stringent version of Predotar, and according to Psort, MetB might be a nuclear protein (Table 11). Calculated physicochemical properties of the deduced MetA and MetB proteins are given in Table 11.

To experimentally investigate the potential of the MetA and MetB N-termini to function as mitochondrial transit peptides, nucleotide sequences encoding the 64 or 57 N-terminal amino acids of MetA or MetB respectively were fused in-frame to the green fluorescent protein (GFP) coding sequence (see II.8.1). Tobacco protoplasts were transformed with the MetA- and MetB-GFP fusion constructs, and transient expression of the fusion proteins was monitored using fluorescence microscopy. Two plasmids encoding mitochondrial CoxIV-GFP and plastidial RecA-GFP fusion protein were used for reference transformations of tobacco protoplasts.

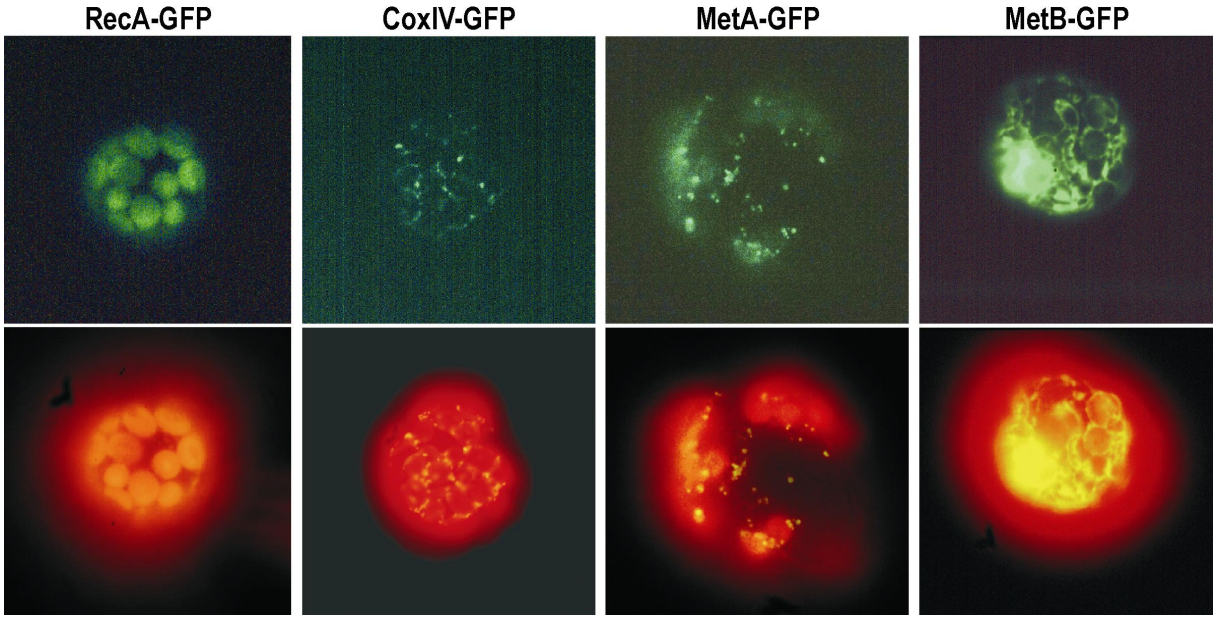



**Figure 16: Transient expression of GFP fusion proteins in tobacco protoplasts.** The MetA and MetB gene fragments encoding putative transit peptides were inserted into plasmid pOL-GFPS65C (Peeters, et al., 2000) to generate vectors driving the expression of MetA-GFP- and MetB-GFP. The control constructs encoding mitochondrial CoxIV-GFP and plastidial RecA-GFP (Peeters, et al., 2000) were kindly provided by I. Small (INRA CNRS, Evry, France). Images were taken by epifluorescence microscopy using a GFP filter (top panels) or a FITC filter set (bottom panels).

Protoplasts expressing MetA-GFP displayed green fluorescence of small structures resembling the fluorescent mitochondria of protoplasts synthesizing CoxIV-GFP (Figure 16), substantiating a mitochondrial localization of MetA. MetB-GFP fluorescence, on the other

hand, for the most part enveloped a large round structure, which likely corresponded to the nucleus, and was moreover distributed over the surfaces of cell organelles such as chloroplasts (as inferred from red chlorophyll autofluorescence in Figure 16, bottom panels). GFP distribution in MetB-GFP-expressing protoplasts pointed to a cytoplasmic localization of the fusion protein, indicating that MetB does not possess a mitochondrial (or plastidial) transit peptide. According to import experiments and targeting predictions (Figure 16 and Table 11), MetB may be a cytoplasmic or nuclear protein. By fusing GFP to only the N-terminal portion of MetB, possible nuclear targeting signals may have been removed.

### III.2.3 Phylogenetic analysis of plant, fungal and animal rRNA dimethylase-like proteins

In order to assess the phylogenetic relationships of Arabidopsis MetA, MetB, and of the plastidial methyltransferase Pfc1 to established mitochondrial transcription factors such as yeast and animal mtTFBs and to other rRNA dimethylases such as *E. coli* KsgA (see I.3.3.1), the MetA, MetB and Pfc1 sequences were compared to available mtTFB sequences and to a set of sequences of characterized and predicted rRNA dimethylases. The latter included all rRNA dimethylase-like ORFs that could be retrieved from the fully sequenced genomes of Arabidopsis, *Populus trichocarpa*, *O. sativa* and *H. sapiens*, as well as additional rRNA dimethylase sequences available from organisms with characterized mtTFBs. Sequence retrieval was done as described in II.9. For plant sequences, a subcellular targeting prediction was performed using the TargetP, Predotar, Mitoprot, Psort and iPsort algorithms (Annex B). Sequences with highest similarity to Arabidopsis MetA were entirely found to have putative mitochondrial transit peptides, and are thus designated MetA in the phylogenetic tree. Sequences best aligning to Arabidopsis MetB were predicted to be neither plastidial nor mitochondrial and are referred to as MetB. The *P. trichocarpa* sequence designated Pt-Pfc1 was calculated to comprise an N-terminal plastidial transit peptide and show highest similarity to Arabidopsis Pfc1.

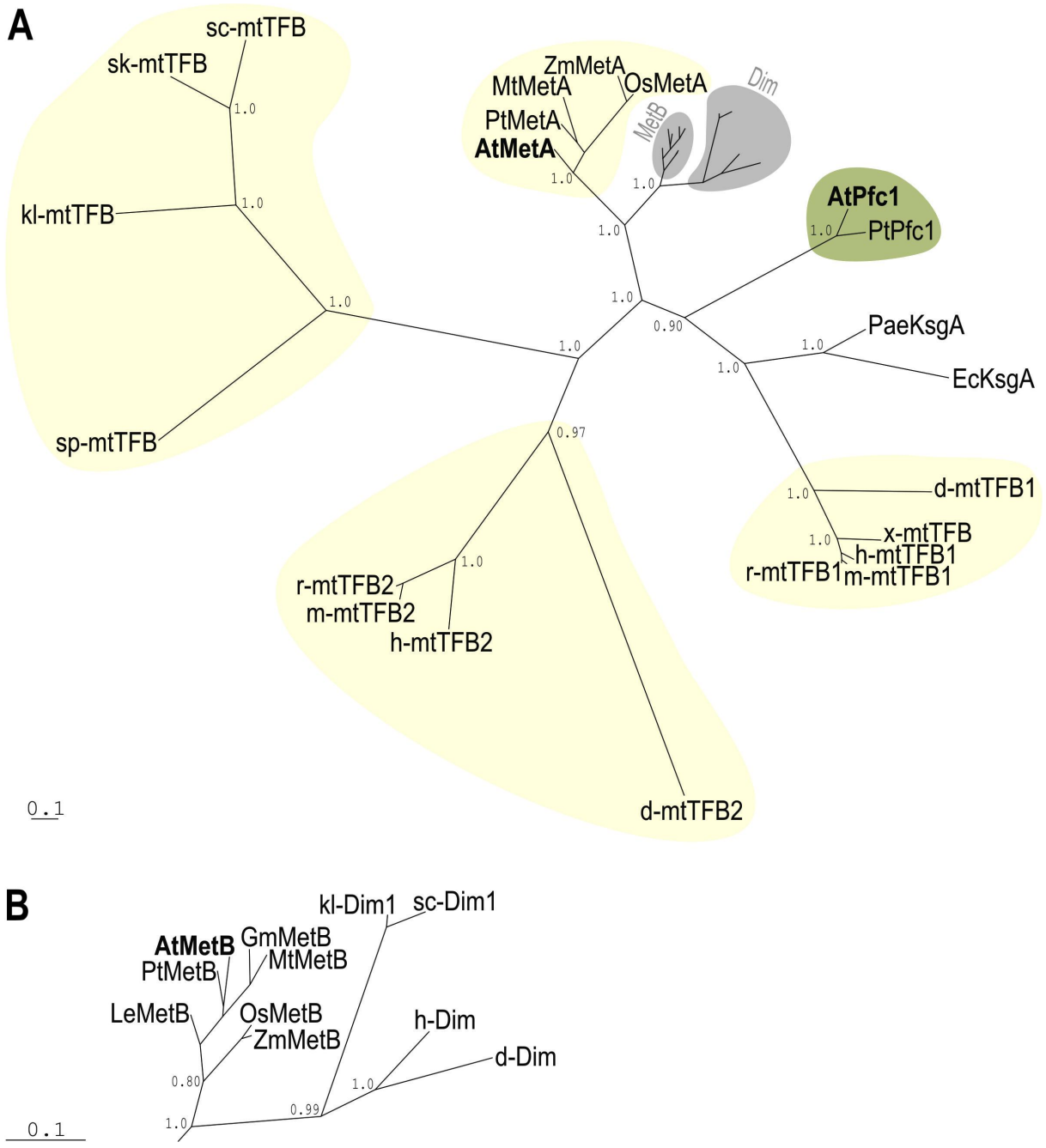



**Figure 17: Phylogeny of mitochondrial transcription factors and small-subunit rRNA dimethylases.** The phylogeny was reconstructed by Bayesian estimation from conserved amino acid sequence sections indicated in the amino acid sequence alignment in Annex B. The clusters comprising plant MetB proteins and yeast and animal Dim enzymes are shaded grey in the comprehensive phylogram under (**A**) and are shown in detail under (**B**). Clades shaded pale yellow comprise mitochondrial or predicted mitochondrial proteins; plastidial components are highlighted green. Prefixes of designations of plant MetA-like, MetB-like and Pfc1-like proteins refer to *Arabidopsis thaliana* (At), *Glycine max*. (Gm), *Lycopersicon esculentum* (Le), *Medicago truncatula* (Mt), *Oryza sativa* (Os), *Populus trichocarpa* (Pt), *Zea mays* (Zm). Fungal and animal protein names adhere to the nomenclature introduced in I.3.3; prefixes refer to *Kluyveromyces lactis* (kl) *Saccharomyces cerevisiae* (sc), *Saccharomyces kluyveri* (sk), *Schizosaccharomyces pombe* (sp), *Drosophila melanogaster* (d), *Homo sapiens* (h), *Mus musculus* (m), *Rattus norvegicus* (r), *Xenopus laevis* (x). EcKsga and PaeKsgA are 16 S rRNA dimethylases from *Escherichia coli* and *Pseudomonas aeruginosa* respectively. Numbers at branching points are posterior branch support values. Branch lengths correspond to the number of inferred amino acid changes per position, as indicated by scale bars.

Amino acid sequences were compared using the Multalin algorithm (Corpet, 1988), and the alignment was refined according to a structure-based alignment generated for fungal mtTFBs and the *Bacillus subtilis* rRNA dimethylase ErmC' (Schubot, et al., 2001); see Annex A for the alignment). Based on the alignment sections that are indicated in Annex A, a Bayesian phylogenetic tree was derived (Figure 17). Essentially the same tree topology was obtained employing maximum likelihood or maximum parsimony analysis (data not shown). Arabidopsis MetB and its plant orthologues appear to be most closely related to a group of rRNA dimethylases containing the yeast nucleolar 18S rRNA dimethylase Dim1. They may therefore represent nuclear or cytoplasmic enzymes, which would be consistent with computational predictions of the subcellular localization of these proteins and with GFP import experiments conducted for Arabidopsis MetB. A sister group to plant, fungal and animal MetB/Dim1-like methyltransferases is formed by predicted mitochondrial rRNA dimethylases of plants including Arabidopsis MetA. Notably, amino acid sequence similarities of Arabidopsis MetA and MetB to yeast Dim1 are 54% and 68% respectively, and considerably exceed similarities to mtTFBs (compare III.2.1). This is contrasted by similarities of sc-mtTFB, h-mtTFB1 and h-mtTFB2 to yeast Dim1 of only 36%, 35% and 28%. Arabidopsis Pfc1 and its poplar orthologue compose a distinct group apart from the Dim1/MetA cluster and from three other well-separated groups formed by animal mtTFB1s, animal mtTFB2s, and the highly diverse fungal mtTFBs. The phylogram shows that plant mitochondrial rRNA dimethylases are decidedly more closely related to nuclear/cytoplasmic enzymes of this type than to fungal and animal mtTFBs. In the absence of any other Arabidopsis mtTFB candidates it was decided to further characterize MetA as a potential cofactor of mitochondrial transcription, and to take along the putative Dim1 orthologue MetB for control experiments.

### III.2.4 Non-specific DNA binding by recombinant MetA

Yeast mtTFB as well as human mtTFB1 have been described previously to bind to mtDNA sequences in a non-specific manner (McCulloch, et al., 2002; Riemen and Michaelis, 1993). Recombinant MetA was prepared in order to assay the protein for a similar DNA-binding activity, and to moreover test the protein *in vitro* for a possible function as cofactor of mitochondrial transcription in Arabidopsis. Using the pPROTet.E expression vector (Clontech), MetA was engineered for expression in *E. coli* as fusion protein carrying an N-terminal hexahistidine tag but lacking the 26 N-terminal amino acids of the methyltransferase, which correspond to the predicted transit peptide and may be expected to not be part of the

mature, functional protein. MetB lacking the 17 N-terminal amino acids was in the same way prepared for expression in *E. coli* so as to have a non-mitochondrial rRNA dimethylase-like protein available for control experiments.

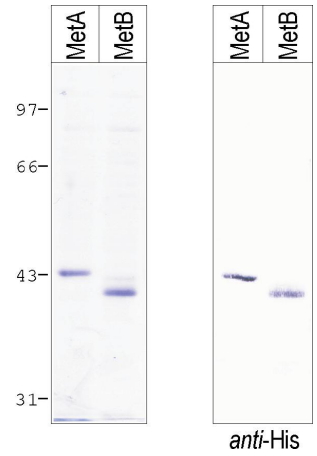



**Figure 18: Purification of recombinant MetA and MetB.** MetA and MetB lacking the predicted transitpeptides (Table 11) and fused to N-terminal hexahistidine tags were expressed in *E. coli*, purified over $Ni^{2+}$-NTA agarose and analyzed by SDS-PAGE followed by Coomassie-blue staining of the gel (left panel). Samples were run alongside a molecular weight marker; sizes are indicated in kDa (marker lane not displayed). Proteins of the expected sizes were found to be enriched, and their identity was confirmed by Western blotting and immunolabelling with an anti-polyhistidine antibody (right panel).

Following expression of the recombinant proteins in the bacterial host, soluble MetA and MetB were enriched from *E. coli* extracts through a $Ni^{2+}$-NTA agarose purification step (see II.5.3.3). Figure 18 displays an SDS-PAGE analysis of the purified proteins. Two major bands migrated as expected for MetA and MetB, and were confirmed to correspond to the recombinant proteins by immunolabelling with an anti-polyhistidine antibody (Figure 18).

Recombinant MetA was tested for DNA-binding activity in an electrophoretic mobility shift assay as described previously for human mtTFB1 (McCulloch, et al., 2002). Two different double-stranded mtDNA fragments that contained either P*atp9*-239 or P*atp9*-295, each representing frequent mitochondrial promoter types in Arabidopsis, were radiolabelled and supplied as target DNA in the binding assay. Binding reactions were subsequently characterized by native PAGE (Figure 19). The addition of MetA to both DNA fragments lead to a mobility shift of the labelled DNA (Figure 19, left panel). This effect was abolished when minor amounts of the non-specific competitor polynucleotide poly(dI-dC) were present in the binding reaction, indicating that the observed DNA binding by MetA is not DNA sequence-specific. Addition of an unlabelled competing mtDNA fragment that did not contain a mitochondrial promoter sequence similarly eliminated the band shift (data not shown). DNA binding by MetB was indistinguishable from the DNA-binding activity of MetA (Figure 19,

right panel). To ensure that the observed DNA-protein complex formation was indeed due to MetA or MetB rather than caused by residual factors from the *E. coli* expression host, *E. coli* extracts from cells containing the empty expression vector pPROTet.E were subjected to the $Ni^{2+}$-NTA agarose purification procedure and subsequently assayed for DNA binding. No band shifts were observed with these fractions, indicating that the DNA-binding activity of the MetA and MetB preparations is indeed due to the recombinant proteins (data not shown). Not only MetA but also MetB behaved like human mtTFB1 in the DNA-binding assay (compare McCulloch, et al., 2002). The proteins were further characterized in *in vitro* transcription studies (see III.4).

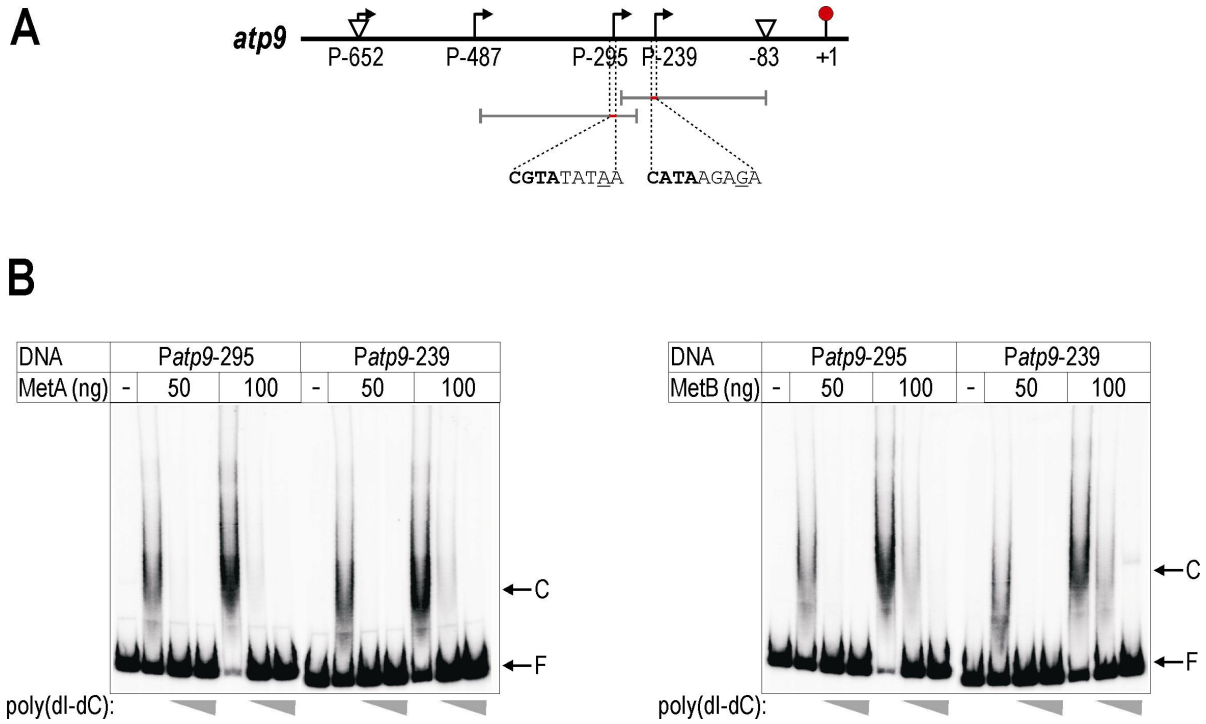



**Figure 19: Gel mobility shift competition assay showing non-specific DNA binding by MetA and MetB.** (**A**) Linear diagram of the *atp9* upstream region. Grey lines mark the positions of the two mtDNA fragments used as probes in the gel mobility shift assay; parts of the comprised promoter sequences are indicated. Positions of transcriptional starts and processing sites are given as negative distances (in base pairs) from the translational start equalling position +1, which is highlighted by a filled circle; other symbols as in Figure 12. (**B**) PAGE analysis of DNA binding by MetA and MetB. Binding reactions were set up with 50 or 100 ng of recombinant MetA (left panel) or MetB (right panel) and with 5' end-labelled mtDNA fragments containing the P*atp9*-295 or the P*atp9*-239 promoter as denoted above the autoradiographs. The addition of increasing amounts (5 and 10 ng) of the non-specific competitor poly(dI-dC) to binding reactions is indicated below. Signals corresponding to DNA-protein complexes (C) and unbound DNA (F) are marked.

## III.3 Expression of the Arabidopsis phage-type RNA polymerases RpoTm and RpoTmp in *E. coli*

The nuclear genes *RpoTm* and *RpoTmp* in Arabidopsis encode a mitochondrial RNA polymerase (RpoTm) and a transcriptase that is imported into both mitochondria and plastids (RpoTmp; see I.3.2.1 and Figure 4; Hedtke, et al., 1997; Hedtke, et al., 2000; Hedtke, et al., 1999). Recombinant RpoTm and RpoTmp were previously shown to non-specifically transcribe DNA *in vitro* (Hedtke, et al., 2000; Kühn, 2001). Transcription of mitochondrial genes by these RNA polymerases, which obligates mitochondrial promoter recognition, has so far not been demonstrated. The present study aims at reconstituting a mitochondrial *in vitro* transcription system from recombinant RpoTm and RpoTmp that is able to accurately initiate transcription at mitochondrial promoters from Arabidopsis, which have been determined here. Therefore, different strategies for the preparation of recombinant RpoT enzymes were evaluated.

*In vitro* transcription assays were formerly done with recombinant RpoTm and RpoTmp lacking the predicted transit peptides and fused N-terminally to thioredoxin (Trx) through expression from the pBAD/Thio vector (Invitrogen) (Kühn, 2001). The Trx tag considerably increased RpoT stability and accumulation in the expression host *E. coli* but might in transcription assays be unfavourable to specific RpoT interactions with promoter sequences or transcriptional cofactors. Proteolytic Trx removal by means of an enterokinase cleavage site separating the Trx and RpoT domains of the recombinant enzymes was previously observed to be associated with partial degradation of the mature protein (Kühn, 2001); D. Stern, BTI, Cornell University, Ithaca, NY, USA, personal communication). To facilitate purification of both Trx-tagged and enterokinase-processed RpoT enzymes, a hexahistidine tag was inserted immediately after the enterokinase processing site (see II.5.1). Figure 20 displays the SDS-PAGE analysis of Trx-hexahistidine-tagged RpoTm and RpoTmp expressed in *E. coli* and enriched from bacterial extracts via a $Ni^{2+}$-NTA agarose.

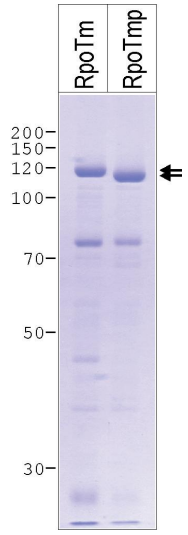



**Figure 20: Purification of recombinant RpoTm and RpoTmp.** RpoTm and RpoTmp lacking the predicted transitpeptides of 42 and 104 amino acids respectively (Kühn, 2001) and fused to an N-terminal Trx-hexahistidine tag were expressed in E. coli (see II.5.2), purified over $Ni^{2+}$-NTA agarose (see II.5.3.1), and analyzed by SDS-PAGE followed by Coomassie-blue staining of the gel. Samples were run alongside a molecular weight marker; sizes are indicated in kDa (marker lane not displayed). Proteins corresponding in size to recombinant RpoTm and RpoTmp [117 and 116 kDa according to ProtParam (http://expasy.cbr.nrc.ca/tools/protparam.html)] were found to be enriched, and their identity was confirmed by immunolabelling (Figure 21).

$Ni^{2+}$-NTA agarose-purified proteins were subjected to the proteolytic removal of the Trx domain by enterokinase (Figure 21). Reaction conditions allowing for the complete processing of Trx-fusion proteins resulted in the enhanced degradation of processed enzymes. Therefore, reactions were optimized to conditions that yielded higher levels of processed hexahistidine-tagged RpoT enzyme than residual undigested protein and only minor amounts of degraded enzyme (Figure 21). Attempts to remove degradation products by a second $Ni^{2+}$-NTA agarose purification step and to further enrich the correctly processed RpoT enzymes resulted in a significant loss of recombinant protein and were therefore not proceeded with.

A strategy to synthesize untagged RpoTm and RpoTmp made use of the pPROTet.E vector employed successfully for MetA and MetB expression. Following removal of the hexahistidine coding sequence from the plasmid, sequences encoding transit peptide-free RpoTm and RpoTmp were inserted (see II.5.1). Figure 22 shows the expression of untagged RpoTm and RpoTmp from these plasmids, which was considerably less stable than expression of Trx-fusion proteins and not visible in *E. coli* extracts on Coomassie-stained gels (not shown). Minor amounts of RpoTm and RpoTmp were detected by immunolabelling with an antibody raised against the RpoTm C-terminus (Figure 22). Purification of these proteins and, most importantly, their separation from the bacterial host RNA polymerase activity (compare Kühn, 2001), may require a series of chromatography steps.

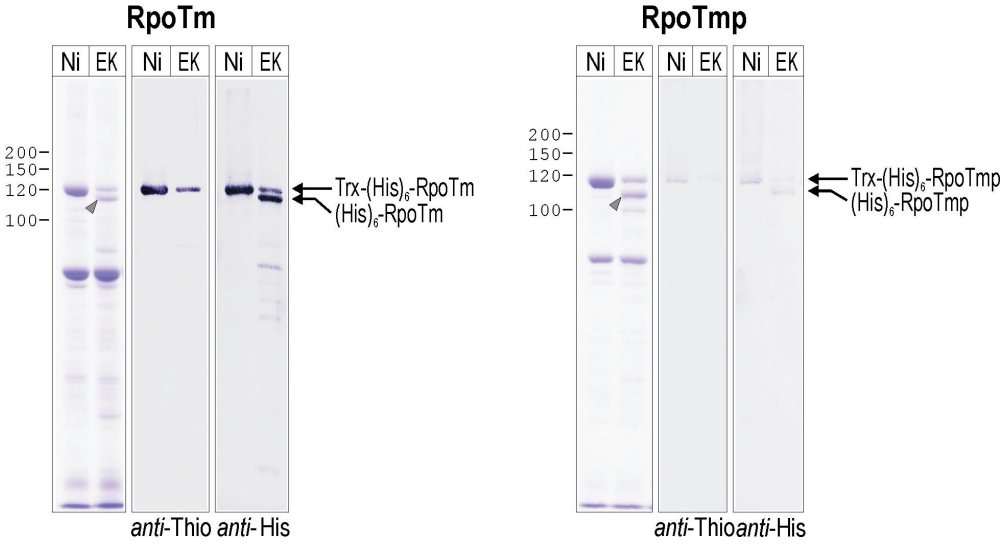



**Figure 21: Proteolytic removal of the Trx domain from Trx-$(His)_6$-tagged RpoTm and RpoTmp.** Enterokinase digests were set up as described in II.5.3.2. $Ni^{2+}$-NTA agarose-purified RpoTm and RpoTmp (lanes designated Ni) and enterokinaseprocessed samples (lanes labelled EK) were analyzed by SDS-PAGE followed by Coomassie-blue staining of gels (left panels), and by Western blotting and immunolabelling of proteins with a thioredoxin antibody (anti-Thio) detecting unprocessed enzymes (middle panels) and a polyhistidine antibody (anti-His) detecting both unprocessed and processed enzymes (right panels). Bands corresponding to Trx-tagged and cleaved Trx-free enzymes are indicated by arrows and specified accordingly. Grey arrowheads mark the correct processing products in Coomassie-stained gels.

Expression of soluble recombinant human mitochondrial RNA polymerase is greatly facilitated by coexpression of h-mtTFB1 or h-mtTFB2 with the enzyme, which results in the formation of stable soluble heterodimers that can be purified by making use of an N-terminal polyhistidine tag attached to h-mtTFB1/2 (Falkenberg, et al., 2002). In the course of the present study, coexpression of untagged RpoTm and RpoTmp with hexahistidine-tagged MetA from an engineered pPROTet.E vector containing two protein expression modules was performed in *E. coli*. The presence of MetA did however not stabilize RpoTm or RpoTmp, as judged from the unimproved synthesis of full-length RpoT and an occasional negative correlation of RpoT and MetA expression (data not shown).

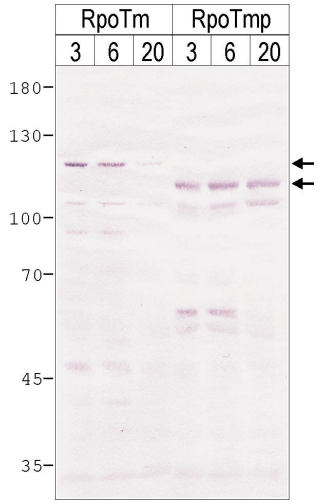



**Figure 22: Expression of untagged RpoTm and RpoTmp in *E. coli*.** RpoTm and RpoTmp lacking the predicted transitpeptides were expressed in *E. coli* from a modified pPROTet.E vector encoding no hexahistidine tag. Following induction of recombinant protein expression, cells were grown for 20 hours at 18°C (see II.5.2) and sampled after 3, 6 and 20 hours. Total cell protein was elecrophoresed by SDS-PAGE and analyzed by Western blotting and immunolabelling with the antibody anti-$Y_c$ detecting both RpoTm and RpoTmp. Sampling times are indicated above individual lanes. Samples were run alongside a molecular weight marker; sizes are given in kDa (marker lane not displayed). Arrows denote bands corresponding to full-length RpoTm and RpoTmp (as opposed to fragments resulting from degradation).

The unsatisfactory results of proteolytic processing of Trx-$(His)_6$-tagged RpoTm and RpoTmp and the insufficient expression of untagged RNA polymerases in *E. coli* prompted Trx-$(His)_6$-tagged RpoTm and RpoTmp to be employed for routine enzyme preparation and *in vitro* transcription experiments.

## III.4 *In vitro* transcription studies of Arabidopsis RpoTm and RpoTmp

### III.4.1 Development of an Arabidopsis *in vitro* transcription system

Chapter III.1 describes the mapping of transcription initiation sites in the mitochondrial genome of Arabidopsis. The knowledge of mtDNA sequences that are recognized as promoters by the transcription machinery *in vivo* enabled an *in vitro* system initiating transcription at Arabidopsis mitochondrial promoters to be set up. To study transcription from mitochondrial promoters *in vitro*, DNA templates were constructed by inserting promoter regions of the Arabidopsis mtDNA into pKL23 (Liere and Maliga, 1999) upstream of the two bacterial ρ-independent terminator sequences *hisa* and *thra* (Barnes and Tuley, 1983; Gardner, 1982) that are present in pKL23 (Figure 23A, Figure 25A, and Figure 27). RNA synthesis driven by a plant organellar phage-type RNA polymerase has been described earlier to efficiently stop at *hisa* and *thra* (Liere and Maliga, 1999). When providing a circular pKL23 derivative as template in run-off experiments, transcription initiated at the introduced promoters should thus be terminated at *hisa* and/or *thra* and/or at downstream cleavage sites on an *Eco*RI- or *Xho*I-linearized plasmid (compare Figure 23A, Figure 25A, and Figure 27), thereby generating RNA products of distinct lengths.

Recombinant Trx-$(His)_6$-tagged RpoTm and RpoTmp were assayed for transcription initiation at several mitochondrial promoters in the presence or absence of the mitochondrial mtTFB homologue MetA. In order to discern non-specific effects of MetA addition, control reactions were set up with the presumably non-mitochondrial MetA homologue MetB. The experimental design essentially followed that described by Falkenberg et al. (2004) for a human mitochondrial *in vitro* transcription system reconstituted from individual recombinant components (see Materials and Methods). Recombinant proteins were prepared for application in *in vitro* transcription assays as described in sections III.3 (RpoT) and III.2.4 (Met).

Plasmid pKL23-*atp6-1*-A containing the promoters P*atp6-1*-156 and P*atp6-1*-200 in tandem was selected as DNA template for initial *in vitro* transcription experiments (Figure 23A; for promoter sequences, see Table 9). Both promoters give rise to fairly abundant *in vitro*-cappable 5' ends (compare Figure 12) and were thus considered efficient *in vivo* promoters. Moreover, they differ in their architecture, thus making pKL23-*atp6-1*-A an ideal template for preliminary studies investigating the experimental conditions for correct promoter utilization *in vitro*.

### III.4.2 *In vitro* transcription from the mitochondrial promoters P*atp6-1*-200, P*trnM*-98 and P*rrn26*-893 by RpoTm

core RNA polymerase. Unlike transcriptionally active preparations from plant mitochondria (Binder, et al., 1995; Hanic-Joyce and Gray, 1991; Rapp and Stern, 1992), RpoTm did not specifically transcribe linear DNA from P*atp6-1*-156 or P*atp6-1*-200 (data not shown, compare Figure 28). Based on a study reporting that a phage-type RNA polymerase activity isolated from tobacco plastids was dependent on supercoiled DNA as a template *in vitro* (Liere and Maliga, 1999), and on previous observations that recombinant RpoTm and RpoTmp were considerably more active in the transcription of supercoiled compared to linear DNA (Kühn, 2001), the experiments were repeated using a circular, negatively supercoiled plasmid template.

Transcription of pKL23-*atp6-1*-A by RpoTm produced three major discrete RNA products of apparent lengths of approximately 300 and 370 nucleotides (Figure 23B; major transcripts are indicated by black arrows). Following their 5' end characterization, these products could be attributed entirely to transcription initiation at P*atp6-1*-200 (see below). While the upper band resulted from transcription termination at *thra*, termination at *hisa* produced an RNA migrating as a double band. Since for all plasmid templates used in subsequent experiments, transcripts ending at *hisa* appeared as double bands, the latter presumably resulted from secondary structure formation at the transcript 3' end or termination at two different adjacent sites at *hisa*. The high-molecular-weight signals that are visible at the top of the autoradiograph may be attributed to transcripts initiated non-specifically on the plasmid template at sequences other than the *atp6-1* promoters and to RNAs not terminated at *hisa* and *thra*. Non-specific high-molecular-weight transcripts were likewise seen in *in vitro* transcription studies with mitochondrial extracts (Binder, et al., 1995; Hanic-Joyce and Gray, 1991; Rapp and Stern, 1992). The addition of equimolar amounts of MetA to RpoTm appeared to have no effect on RpoTm-driven transcription. No transcripts were made in reactions containing MetA or MetB but not RpoTm. It is thus unlikely that the transcripts seen in Figure 23B are due to an RNA polymerase activity from *E. coli* exploited as RpoTm expression host, as RpoTm, MetA, MetB preparations were made following essentially the same protocol and may be expected to contain similar residual *E. coli* contaminants. Subsequent *in vitro* experiments revealing non-identical transcription activities of RpoTm and RpoTmp (see below) supported that the observed *in vitro* RNA synthesis was indeed due to the recombinant enzymes.

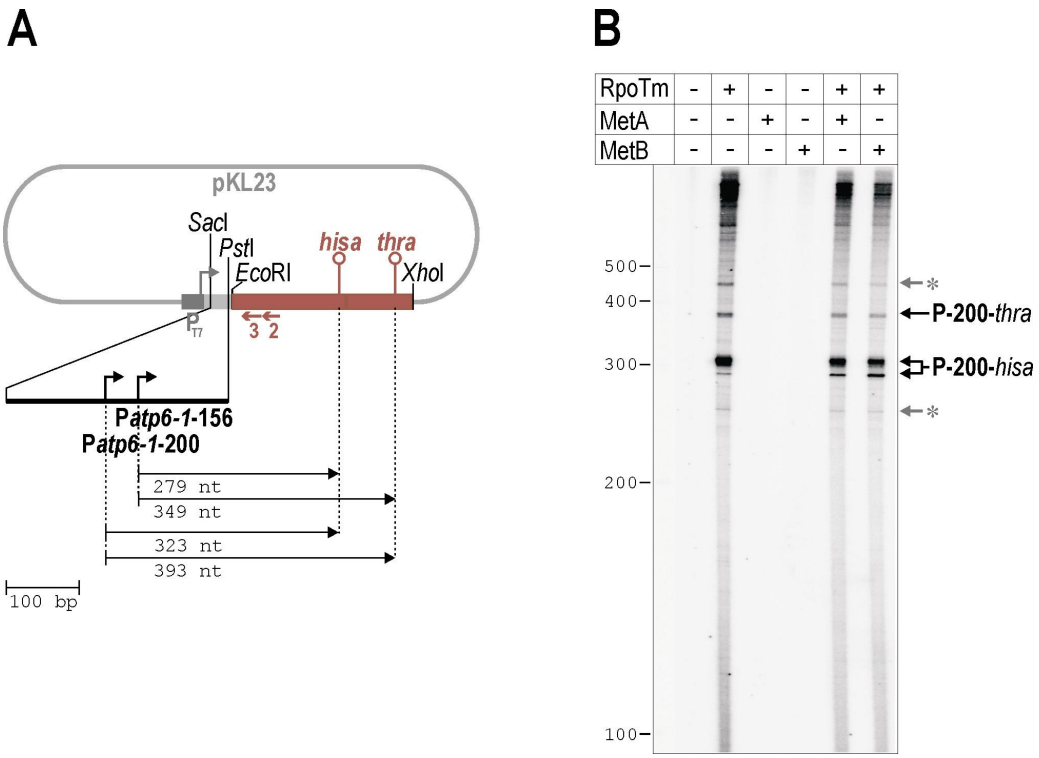



**Figure 23: *In vitro* run-off transcription of pKL23-atp6-1-A by RpoTm initiates at Patp6-1-200.** (**A**) pKL23-atp6-1-A was constructed by inserting a 300-bp fragment of the Arabidopsis mtDNA containing the promoters P*atp6-1*-156 and P*atp6-1*-200 (compare Figure 31) into pKL23 via the *Sac*I and *Pst*I restriction sites. The positions of a T7 promoter ($P_{T7}$) represented by a dark grey bar and of the bacterial attenuators *hisa* and *thra* symbolized by red bars are indicated. Red open circles denote sites of transcription termination within *hisa* and *thra*; bent arrows mark start points of transcription at P*atp6-1*-156, P*atp6-1*-200, and $P_{T7}$. Only the plasmid region between $P_{T7}$ and the *Xho*I cleavage site, including the mtDNA insert, is drawn to scale. Run-off products expected from initiation at P*atp6-1*-156 and P*atp6-1*-200 and termination at *hisa* and *thra* are indicated by horizontal black arrows labelled with the respective RNA length. Red arrows and numbers mark the positions of primers P2hisa (2) and P3hisa (3) employed for transcript 5' end mapping. (**B**) RpoTm was assayed for promoter-specific transcription of supercoiled pKL23-atp6-1-A in the presence or absence of MetA or MetB as described in section II.7.2. [□-$^{32}$P]-UTP-labelled RNA products were electrophoresed in a 5% sequencing gel alongside an RNA size standard; sizes are given in nucleotides. *In vitro* transcription reactions were supplemented with recombinant proteins (400 fmol each) as indicated above individual lanes. Major discrete RNA products are indicated by black arrows and, after transcript 5' end mapping (Figure 24), were attributed to transcription initiation at P*atp6-1*-200 followed by termination at *hisa* (signals labelled P-200-*hisa*) and *thra* (signal P-200-*thra*) as detailed in the text. Minor signals indicated by grey arrows labelled with asterisks may be due to differently migrating major products.

From a comparison of expected and apparent transcript lengths (Figure 23), the discrete *in vitro* transcription products indicated in Figure 23B could be assigned neither to transcription initiation at P*atp6-1*-156 nor to initiation at P*atp6-1*-200. Discrepancies between expected and apparent sizes were observed for nearly all *in vitro*-synthesized transcripts and may be due to an altered migration behaviour of these products compared to the RNA size marker. In order to identify the 5' termini of *in vitro*-synthesized transcripts, 5'-RACE was performed on RNA products made by RpoTm essentially as described in section III.1.1 (Figure 24, lane +L). As a control, *in vitro* transcription products not ligated to the RNA oligonucleotide were subjected to RT-PCR (Figure 24, lane -L), thereby allowing to distinguish RNA-derived PCR products

from signals resulting from non-specific amplification of of sequences from the carried-over DNA template.

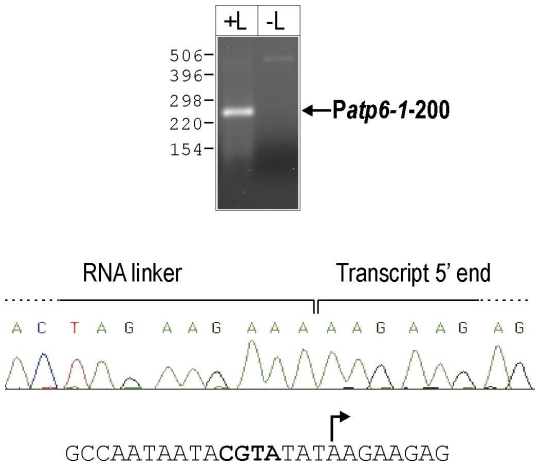



**Figure 24: 5'-RACE analysis of pKL23-atp6-1-A-derived transcripts synthesized in vitro by RpoTm.** 5'-RACE was performed on RNA linker-ligated transcripts (lane +L) and, as a control, on non-ligated transcripts (-L). PCR products were separated on an agarose gel alongside a molecular weight marker; sizes are given in base pairs (marker lane not displayed). The signal that corresponds to transcript 5' ends mapping to P*atp6-1*-200 is indicated. The chromatogram below displays the sequence at the ligation site of a typical cloned 5'-RACE product; RNA linker and transcript portions of the sequence are indicated. The mtDNA sequence at P*atp6-1*-200 is displayed below; the bent arrow indicates the *in vivo* transcription initiation site.

Electrophoresis of 5'-RACE products showed a single band, indicating that the multiple transcripts synthesized from pKL23-*atp6-1*-A by RpoTm are due to transcription initiation at only one of the two promoters (Figure 24). Sequencing of the cloned PCR product revealed that the transcripts had a 5' end identical to that of transcripts initiated at P*atp6-1*-200 *in vivo*, allowing to attribute the *in vitro*-synthesized RNAs to correct transcription initiation at this promoter (Figure 24). On the other hand, *in vitro* transcription and 5'-RACE experiments together provided no evidence for transcripts initiated specifically at P*atp6-1*-156. The minor signals indicated by grey arrows and marked with asterisks in Figure 23B were assumed to correspond to differently migrating major transcripts rather than additional defined RNA 5' ends, as no discrete 5'-RACE products other than those resulting from initiation at P*atp6-1*-200 were detected.

Additional *in vitro* transcription experiments investigated the initiation by RpoTm at the promoters P*rrn26*-893 and P*trnM*-98. While the sequence around the transcription initiation site of P*rrn26*-893 is nearly identical to that of P*atp6-1*-156, P*trnM*-98 displays elements of both P*atp6-1*-156 and P*atp6-1*-200 (see Table 9). Supercoiled pKL23-*rrn26* and pKL23-*trnM* gave rise to discrete RNA products (Figure 25), whereas no specific transcripts were seen with the linearized plasmids (data not shown). The transcript 5' termini were found for both templates to accurately reflect those of transcripts generated *in vivo* from P*rrn26*-893 and

P*trnM*-98, respectively (Figure 26). Initiation at P*rrn26*-893 and P*trnM*-98 was, however, considerably less efficient than at P*atp6-1*-200. As observed with pKL23-*atp6-1*-A, MetA had no effect on RpoTm-driven transcription from pKL23-*rrn26* and pKL23-*trnM*.

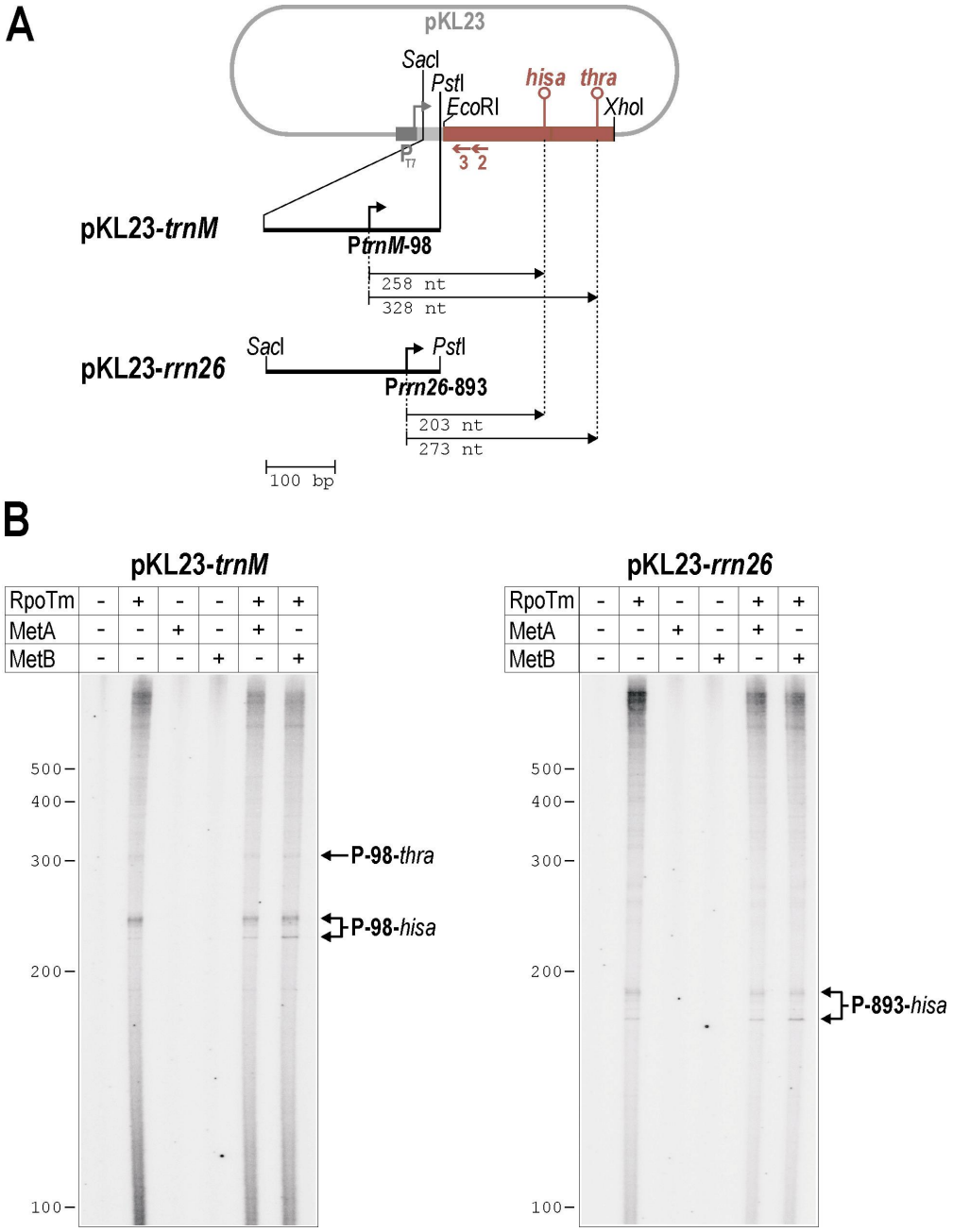



**Figure 25: *In vitro* transcription from P*trnM*-98 and P*rrn26*-893 by RpoTm.** (**A**) pKL23-trnM and pKL23-rrn26 were constructed by inserting fragments of the Arabidopsis mtDNA containing the promoter P*trnM*-98 or P*rrn26*-893 (compare Figure 31) into pKL23 via the *Sac*I and *Pst*I restriction sites. Symbols and illustration of expected run-off products as in Figure 23A. (**B**) RpoTm was assayed for promoter-specific transcription of supercoiled pKL23-trnM or pKL23-rrn26 in the presence or absence of MetA or MetB as indicated, and RNA products were analyzed as described above for pKL23-atp6-1-A. Major discrete RNA products are indicated by arrows and, after transcript 5' end mapping (Figure 26), were attributed to transcription initiation at P*trnM*-98 followed by termination at *hisa* (signals labelled P-98-*hisa*) and *thra* (signal P-98-*thra*), or initiation at P*rrn26*-893 followed by termination at *hisa* (P-893-*hisa*) and *thra* (P-893-*thra*).

The described *in vitro* assays showed RpoTm to specifically initiate transcription at three different mitochondrial promoters, P*atp6-1*-200, P*rrn26*-893 and P*trnM*-98, from supercoiled DNA without the aid of auxiliary factors. The preference of promoters over random start sites varied between different promoter sequences.

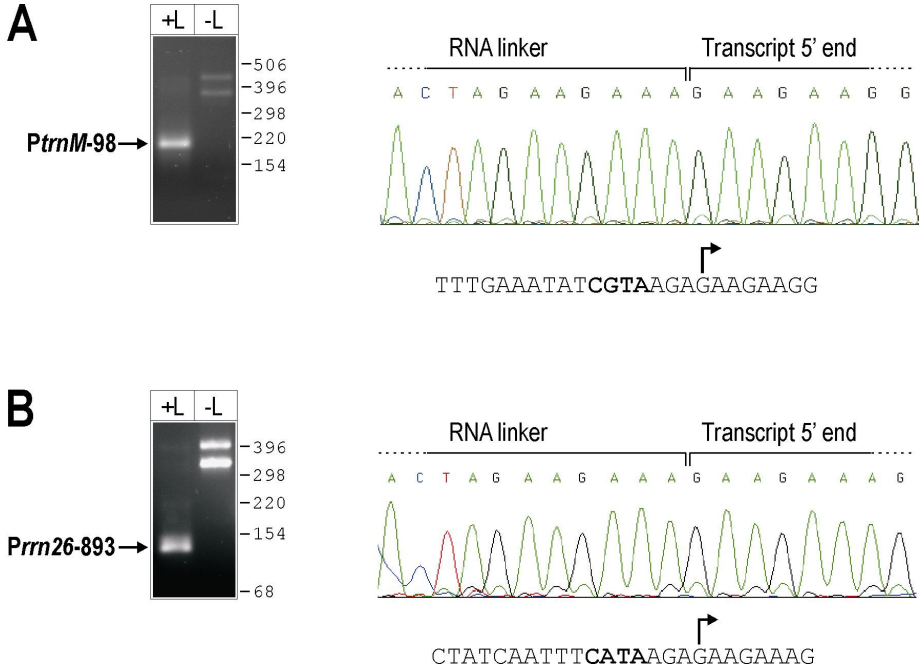



**Figure 26: 5'-RACE analysis of RNAs transcribed from P*trnM*-98 and P*rrn26*-893 by RpoTm *in vitro*.** (**A**) Analysis of pKL23-trnM-derived RNAs. 5'-RACE was performed on RNA linker-ligated transcripts (lane +L) and, as a control, on non-ligated transcripts (-L). PCR products were separated on an agarose gel alongside a molecular weight marker (left); sizes are given in base pairs (marker lane not displayed). The signal corresponding to transcript 5' ends mapping to P*trnM*-98 is indicated. The chromatogram to the right displays the sequence at the ligation site of a typical cloned 5'-RACE product; RNA linker and transcript portions of the sequence are indicated. The mtDNA sequence at P*trnM*-98 is displayed below; the bent arrow indicates the *in vivo* transcription initiation site. (**B**) Analysis of pKL23-rrn26-derived RNAs as described under (A). The 5'-RACE signal corresponding to transcript 5' ends mapping to P*rrn26*-893 is indicated (left); the sequence at P*rrn26*-893 is displayed below the chromatogram of a typical cloned 5'-RACE product (right).

### III.4.3 Comparison of the transcriptional performances of RpoTm and RpoTmp

Arabidopsis mitochondria possess two phage-type RNA polymerases (Hedtke, et al., 1997; Hedtke, et al., 2000); yet no data have been provided so far that would define possible functional differences between RpoTm and RpoTmp. A series of experiments therefore compared the abilities of RpoTm and RpoTmp to initiate transcription *in vitro* at diverse mitochondrial promoter sequences. Figure 27 illustrates the design of plasmid templates that were constructed for this study and supplied as supercoiled or linearized templates (for promoter sequences, see Table 9).

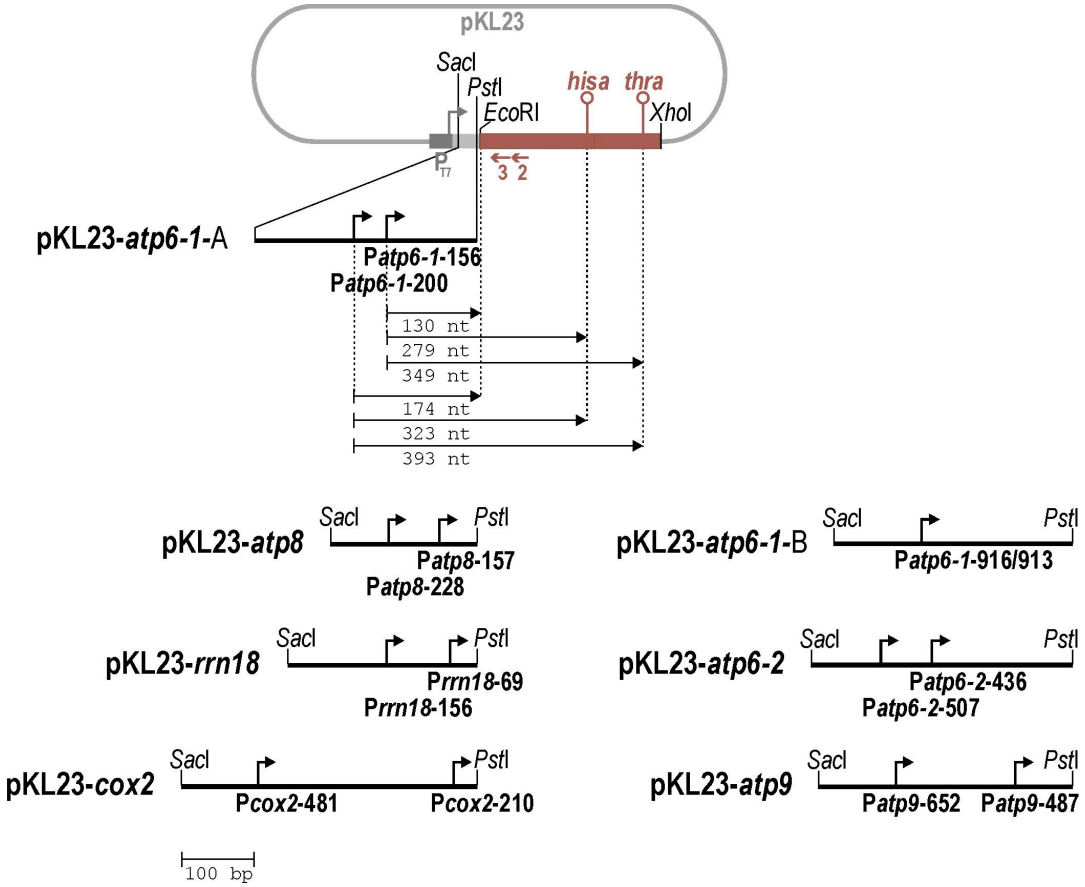



**Figure 27: Maps of *in vitro* transcription templates.** pKL23 derivatives were constructed by inserting Arabidopsis mtDNA fragments containing promoter sequences of the *atp6-1, atp6-2, atp8, atp9, rrn18*, and *cox2* genes (compare Figure 31) into pKL23 via the *Sac*I and *Pst*I restriction sites. RNA molecules expected from transcription initiation at the introduced promoters followed by termination at *hisa* or *thra* or the restriction site used for plasmid linearization are exemplified for pKL23-atp6-1-A. Symbols and illustration of run-off products as in Figure 23A. The following RNAs may be expected from transcription of linearized or supercoiled pKL23 derivatives:

| Template | Promoter | Cleavage site | Expected transcript length due to termination at | | |
|---|---|---|---|---|---|
| | | | cleavage site | *hisa* | *thra* |
| pKL23-*atp6-1*-A | P*atp6-1*-156 | *Eco*RI | 130 nt | 279 nt | 349 nt |
| | P*atp6-1*-200 | | 174 nt | 323 nt | 393 nt |
| pKL23-*atp8* | P*atp8*-157 | *Xho*I | 309 nt | 206 nt | 276 nt |
| | P*atp8*-228/226 | | 380 nt | 277 nt | 347 nt |
| pKL23-*rrn18* | P*rrn18*-69 | *Xho*I | 295 nt | 192 nt | 262 nt |
| | P*rrn18*-156 | | 382 nt | 279 nt | 349 nt |
| pKL23-*cox2* | P*cox2*-210 | *Xho*I | 290 nt | 187 nt | 257 nt |
| | P*cox2*-481 | | 561 nt | 458 nt | 528 nt |
| pKL23-*atp6-1*-B | P*atp6-1*-916/913 | - | - | 369/366 nt | 439/436 nt |
| pKL23-*atp6-2* | P*atp6-2*-436 | - | - | 350 nt | 420 nt |
| | P*atp6-2*-507 | | - | 421 nt | 491 nt |
| pKL23-*atp9* | P*atp9*-487 | - | - | 233 nt | 307 nt |
| | P*atp9*-652 | | - | 398 nt | 468 nt |

*In vitro* transcription of pKL23-*atp6-1*-A confirmed initiation at P*atp6-1*-200 but not P*atp6-1*-156 on supercoiled but not *Eco*RI-cleaved DNA by RpoTm (Figure 28). While non-specific RpoTmp-driven transcription of pKL23-*atp6-1*-A far exceeded RpoTm activity, no specifically initiated RNAs were seen in reactions with RpoTmp. Neither did RpoTmp

recognize P*atp6-1*-156 or P*atp6-1*-200 on the linearized plasmid (data not shown). The presence of MetA appeared to have no effect on the activity or specificity of RpoTmp.

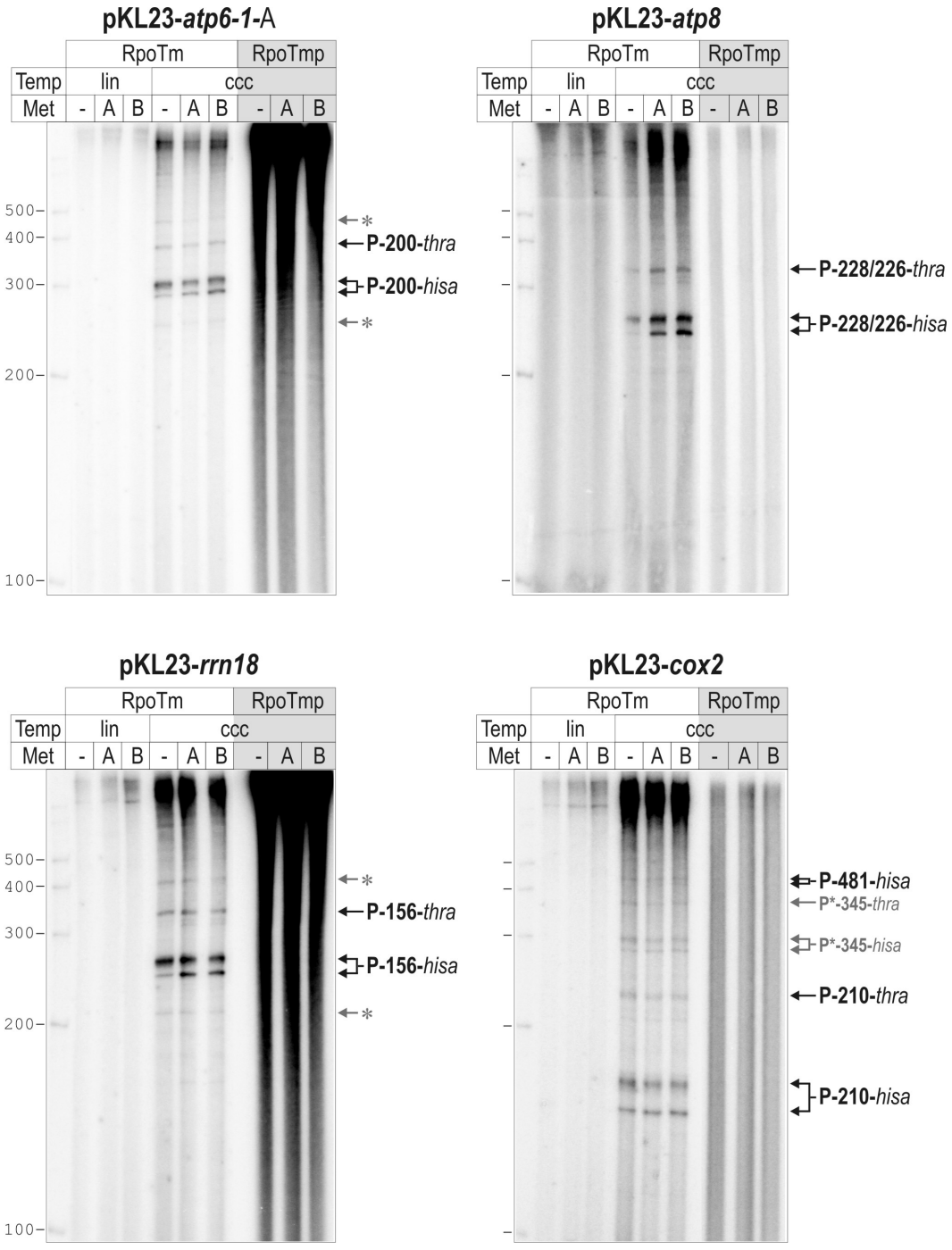



**Figure 28: Run-off transcription from the *atp6-1*, *atp8*, *rrn18* and *cox2* upstream regions by RpoTm and RpoTmp.** RpoTm and RpoTmp were assayed for promoter-specific transcription of pKL23-atp6-1-A, pKL23-atp8, pKL23-rrn18 and pKL23-cox2. In vitro transcription reactions were run with RpoTm or RpoTmp and supercoiled (ccc) or linear (lin) DNA in the presence or absence of MetA (A) or MetB (B) as indicated above lanes, and RNA products were analyzed as described above. Transcripts made by RpoTmp were diluted 1:5 prior to PAGE due to RpoTmp showing a markedly higher activity than RpoTm. Discrete RNA products are specified as in Figure 23. Signals denoted P*-345-*hisa* and P*-345-*thra* are due to non-specific initiation on pKL23-cox2 (see text). Different background signal intensities observed for the same enzyme and template conformation are due to inherent experimental variation.

Exclusively non-specific transcription of linear templates was also observed in all subsequent *in vitro* transcription experiments for both RpoTm and RpoTmp (data not shown for RpoTmp). Modified experimental conditions, such as altered concentrations of monovalent ions or the catalytic $Mg^{2+}$, did not enable RpoTm to initiate RNA synthesis at promoters located on linear templates or at promoters not recognized under the conditions of the standard *in vitro* transcription protocoll (see II.7.2; linear and supercoiled templates were purified using the same purification system). Neither RpoTm nor RpoTmp was found to be influenced in its activity by the addition of different amounts of MetA (only experiments in which equimolar amounts of RpoTm/RpoTmp and MetA were used are displayed).

Of the two mitochondrial promoters residing on pKL23-*atp8*, only P*atp8*-228/226 significantly supported specific transcription initiation by RpoTm while P*atp8*-157, which displays a sequence identical to P*atp6-1*-156 around the *in vivo* transcription start site (see Table 9), was apparently not recognized by RpoTm as a promoter (Figure 28). Enhanced transcription of pKL23-*atp8* by RpoTm in the presence of MetA was considered irrelevant as a similar increase in RpoTm activity was seen following MetB addition. 5'-RACE performed on pKL23-*atp8*-derived RNAs confirmed the major defined transcripts to map to one out of two adenines in P*atp8*-228/226 used as initiating nucleotides by the mitochondrial transcription machinery *in vivo* (compare Table 9 and Table 12). Extensive sequencing of cloned 5'-RACE products identified minor discrete transcript 5' ends mapping to P*atp8*-157 (Figure 29). However, these transcripts were not distinguishable from the background of non-specific products in the autoradiograph (Figure 28).

RpoTm-driven transcription of pKL23-*rrn18* was found to efficiently and accurately initiate at P*rrn18*-156 but not P*rrn18*-69 *in vitro* (Figure 28 and Figure 29). The minor signals indicated by grey arrows and marked with asterisks in Figure 28 were assumed to correspond to differently migrating major transcripts rather than additional defined RNA 5' ends, as they did not give rise to discrete 5'-RACE products (Figure 29). Minor signals at similar distances from major bands were likewise seen for the template pKL23-*atp6-1* (Figure 23). Transcripts with altered migration behaviour may have retained unmelted secondary structures in the denaturing polyacrylamide gel.

Transcription of pKL23-*cox2*, which harbours two mitochondrial promoters, yielded discrete RNA products mapping to three different initiation sites (Figure 28). Two of these sites were through 5'-RACE confirmed to correspond to the *in vivo* initiating nucleotides of P*cox2*-210 and P*cox2*-481 (Figure 29). The P*cox2*-481 sequence additionally gave rise to a transcript 5' end not observed *in vivo* (Table 12). Erratic initiation by RpoTm (signals

indicated by grey arrows in Figure 28) occurred at a sequence that resembles Arabidopsis mitochondrial promoters but had not been detected to function as promoter *in vivo*. The 5'-terminal nucleotide of the non-specific transcript was determined to correspond to position -345 upstream of the beginning of the *cox2* coding sequence on the Arabidopsis mtDNA, and the initiation site was denoted P**cox2*-345 (Figure 28 and Figure 29, and Table 12). Of the plasmids pKL23-*atp8*, pKL23-*rrn18* and pKL23-*cox2*, none supported specific transcription initiation by RpoTmp (Figure 28). Neither did pKL23-*rrn26* or pKL23-*trnM* stimulate specific transcription by RpoTmp (data not shown).

In order to rule out the possibility that the failure of RpoTm to efficiently initiate at P*atp6-1*-156, P*atp9*-239, and P*rrn18*-69 was due to the presence of a second, favoured upstream promoter on the plasmid template (compare Figure 27), additional templates were designed containing the P*atp6-1*-156, P*atp9*-239, or P*rrn18*-69 promoter region but lacking the upstream P*atp6-1*-200, P*atp9*-295, or P*rrn18*-156 sequences, respectively. However, the removal of the latter promoters from *in vitro* transcription templates did not enhance initiation at P*atp6-1*-156, P*atp9*-239, or P*rrn18*-69 (data not shown).

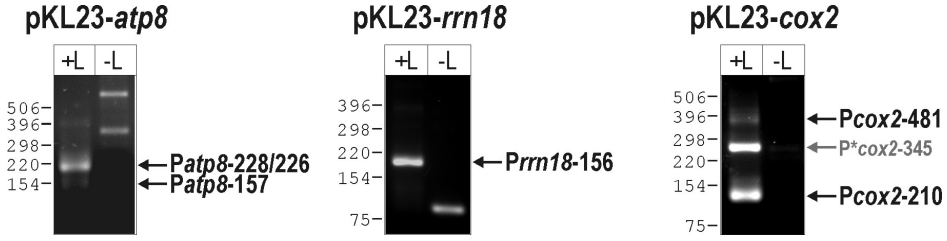



**Figure 29: 5'-RACE analysis of *in vitro*-synthesized transcripts indicated in Figure 28.** 5' end mapping was performed as described (see Figure 24) on RNAs synthesized by RpoTm from supercoiled pKL23-atp8, pKL23-rrn18, and pKL23-cox2 (chromatograms not displayed). Signals are labelled with the corresponding promoter name or initiation site.

In the described *in vitro* assays, specific utilization of mitochondrial promoters as transcription start sites by RpoTm depended on a supercoiled conformation of DNA templates. RpoTmp was unable to specifically initiate transcription, regardless of the template structure. Promoter-specific transcription initiation by the two RNA polymerases was apparently not stimulated by the mtTFB-like protein MetA.

### III.4.4 Transcription initiation by RpoTm and RpoTmp at non-CRTA promoters

In a study by Binder et al. (1995), a transcriptional activity prepared from pea mitochondria specifically and exclusively initiated transcription at promoters essentially conforming to the CRTAAGAGA nonanucleotide consensus derived previously for dicot

mitochondrial promoters (Binder, et al., 1995). Recognition of a deviating mitochondrial promoter by this *in vitro* transcription system was later described by a single report (Kuhn and Binder, 2002). Additional promoters of the Arabidopsis mtDNA that do not possess a CRTA core element were thus tested for their ability to direct transcription by RpoTm or RpoTmp *in vitro*.

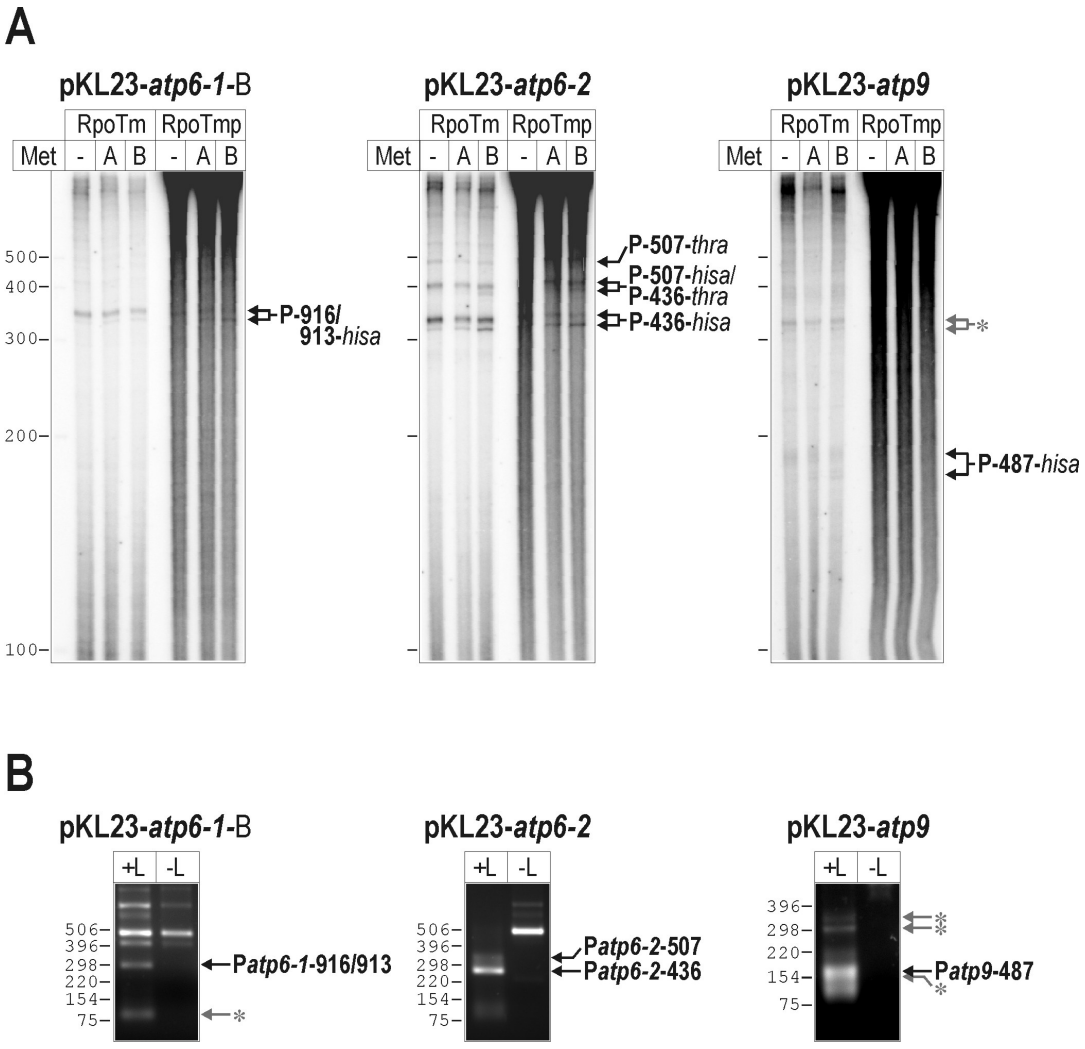



**Figure 30:** ***In vitro*** **transcription initiation at promoters not displaying a CRTA sequence element.** (**A**) RpoTm and RpoTmp were assayed for promoter-specific transcription of supercoiled pKL23-atp6-1-B, pKL23-atp6-2, and pKL23-atp9. In vitro transcription reactions were run with RpoTm or RpoTmp in the presence of MetA (A) or MetB (B) as indicated, and RNA products were analyzed as described above. Transcripts made by RpoTmp were diluted 1:5 prior to PAGE. Discrete RNA products are specified as in Figure 23. (**B**) 5' end mapping was performed as described on RNAs synthesized by RpoTm from pKL23-atp6-1-B, pKL23-atp6-2, and pKL23-atp9 (chromatograms not displayed). Signals are labelled with the corresponding promoter name; asterisks denote non-specific products amplified from RNA linker-ligated transcripts (lane +L). Different background signal intensities observed for the same enzyme and template conformation are due to inherent experimental variation.

While no specific RNAs were synthesized from supercoiled plasmids containing the P*rps3*-1053, P*rps3*-1133, and P*trnM*-574/573 sequences (data not shown), defined transcripts were made from P*atp6-1*-916/913, P*atp6-2*-436, P*atp6-2*-507, and P*atp9*-487 by RpoTm (Figure 30; for promoter sequences, see Table 9). 5'-RACE showed that transcription initiated *in vitro* at P*atp6-1*-916/913 at the two adenine nucleotides defined previously as *in vivo* transcriptional starts (Figure 30 and Table 12). The RNA 5' termini identified for P*atp6-2*-436 and P*atp6-2*-507 did not exactly correspond to the *in vivo* primary 5' ends but mapped to positions one and three nucleotides respectively downstream of the correct start sites (Table 12). Minor discrete RNAs equalling in size those obtained with RpoTm were generated by RpoTmp from pKL23-*atp6-1*-B and pKL23- *atp6-2*. Precise mapping of the 5'-termini of these products was not attempted as they were hardly visible within the background of far more abundant non-specific transcripts, thus rendering the identification of defined RNA 5' termini extremely difficult. The preference of P*atp9*-487 over random initiation sites by RpoTm was very weak. While RpoTm did apparently not recognize P*atp9*-652 as a promoter, additional defined transcripts synthesized from pKL23-*atp9* were seen which mapped to non-specific sites. No specific transcripts were made by RpoTmp from pKL23-*atp9*.

*In vitro* transcription experiments demonstrated RpoTm to recognize diverse promoter sequences, though not all promoters tested, on supercoiled templates without the aid of transcriptional cofactors. In contrast, RpoTmp did not significantly prefer mitochondrial promoters over random initiation sites.

Two mitochondrial proteins from Arabidopsis encoded at loci At1g80270 and At1g15480 (mitochondrial targeting confirmed by B. Hedtke, HU Berlin), which are homologous to the wheat PPR protein p63 implicated in mitochondrial transcription (Ikeda and Gray, 1999); see I.3.3.3), were additionally tested in *in vitro* transcription assays for their potential to modulate the transcriptional performances of RpoTm and RpoTmp. Unlike reported for wheat p63, the Arabidopsis proteins did not stimulate transcription initiation at mitochondrial promoters by RpoTm or RpoTmp (*in vitro* transcription experiments were set up both with and without MetA; data not shown). A transcriptional role of Arabidopsis p63-like proteins was thus not confirmed.

# IV DISCUSSION

## IV.1 Multiple promoters as a common feature of mitochondrial genes in Arabidopsis

### IV.1.1 Identification of transcription initiation sites

Our current knowledge of Arabidopsis mitochondrial promoters is largely based on computational predictions performed on the mitochondrial genome sequence of this plant. While a set of 29 hypothetical mitochondrial promoters has emerged from searching the Arabidopsis mtDNA for motifs known to drive transcription initiation in other dicotyledonous plants (Brennicke, et al., 1999; Dombrowski, et al., 1998), experimental evidence is limited to a single promoter in Arabidopsis mitochondria (Giese, et al., 1996). Here, the architecture and distribution of promoters in the upstream regions of twelve mitochondrial genes in Arabidopsis have been analyzed through mapping transcription start sites by a 5'-RACE technique that has previously been employed successfully for the analysis of bacterial and plastidial transcript 5' termini (Argaman, et al., 2001; Bensing, et al., 1996; Miyagi, et al., 1998; Vogel, et al., 2003).

The method proved a valid tool for mitochondrial 5' end detection, and for 19 out of 30 mapped promoters clearly discriminated between primary and processed transcript ends. However, for selected 5' termini that were later identified through analysis of *in vitro*-capped transcripts to result from transcription initiation, 5'-RACE results did not support transcription initiation at the respective promoters. One example is P*rrn18*-69, for which the corresponding 5'-RACE product was significantly more abundant after amplification from transcripts not treated with TAP. Most likely, the misleading PCR result is due to a highly abundant primary transcript derived from the upstream promoter P*rrn18*-156, which is favoured in the 5'-RACE from TAP-treated transcripts, thereby outcompeting the 5'-RACE product generated from the much less abundant downstream 5' terminus (compare *rrn18* 5'-RACE products in Figure 8 and *in vitro*-capped transcripts in Figure 12).

For a group of primary transcripts of the *atp1*, *atp6-1*, *atp6-2*, *atp8* and *atp9* genes mapping to CGTATATAA or CATAAGAGA motifs, 5'-RACE failed to distinguish primary from 5'-processed transcripts, owing to both types of transcripts mapping to the same initiating nucleotide. The processed transcripts could be derived from a modification of primary transcripts by a phosphatase or pyrophosphatase, or from endonucleolytic cleavage of transcripts initiated at upstream promoters.

Various other mechanisms of transcript 5' end processing are implied from the analysis of mitochondrial RNA 5' termini. Primary transcripts originating from P*cox2*-210 and

P*cox2*-481 appear to be trimmed by nucleolytic activities at several consecutive sites immediately downstream of the primary 5' end, yielding processed transcripts of slightly different lengths. It is possible, however, that the shortened transcripts are not *in vivo* products but artefacts due to instability of specific transcripts during the experimental procedure. Several processed 5' ends such as those mapping to positions -83 upstream of the *atp9* coding sequence and -678 in the cox2 5' region (compare Figure 31) are likely to have been generated through endonucleolytic removal of an RNA fragment from a newly initiated transcript. Endonucleases may be expected to function in mitochondrial transcript maturation in Arabidopsis, as all primary rRNA or tRNA 5' ends defined in this study map to positions upstream of the 5' terminus of the respective mature RNA. While 5'-terminal processing of tRNAs in plant mitochondria has been established to be carried out by RNase P or its homologue, enzymes involved in 5'-end maturation of rRNAs or mRNAs are as yet unidentified (Binder and Brennicke, 2003; Marchfelder, et al., 1996).

### IV.1.2 Promoter architecture

Nearly 50% of the transcriptional starts identified in this study are located within sequences that conform to the nonanucleotide consensus CRTAAGAGA previously suggested for dicot mitochondrial promoters (Binder, et al., 1996) or the motif CGTATATAA defined in this work (Table 9). About as many promoters deviate to a varying extent in the sequence surrounding the transcriptional start and in place of a CRTA core display ATTA or RGTA tetranucleotides that have emerged from this study as frequent promoter elements. 20 out of 30 promoters support initiation at an adenine nucleotide, which complements previous reports on transcription initiation mostly at guanine nucleotides in dicot mitochondria (Fey and Marechal-Drouard, 1999). While purines appear obligatory at positions +1 and +2 with respect to the transcriptional start, they are moderately frequent at positions +3 to +8. Preceding the promoter core is usually a sequence rich in A and T nucleotides, which has been described for several mitochondrial genes in dicots as well as in maize to be important for unimpeded promoter function *in vitro* (Dombrowski, et al., 1999; Rapp, et al., 1993). Moreover, the all-A/T sequence TATATA is seen as an element not only of the CGTATATAA motif but also of deviating promoters such as P*rrn18*-353 and P*atp6-1*-916/913 (Table 9), emphasizing the predisposition of A/T-rich nucleotide sequences to function as promoters. In line with this idea is the observation that of the two promoters P*atp8*-710 and P*trnM*-574/573 showing neither a recognizable core motif nor the TATATA element, the latter is highly rich in A-T base pairs. The ability of mitochondrial sequences

composed entirely of A and T nucleotides to support transcription initiation has been pointed out by (Lupold, et al., 1999), based on their own observation of an all-A/T promoter of the maize *cox2* gene and an earlier report on yeast petite mutants presumably initiating mitochondrial RNA-synthesis at all-A/T sequences (Fangman and Dujon, 1984).

The detection of transcription initiation sites in the Arabidopsis mtDNA was strongly biased towards the identification of promoters displaying a CRTAAGAGA or CGTATATAA motif, as 9 out of 12 genes were selected for the analysis of transcriptional starts owing to the prior observation of at least one of these motifs in their 5' regions. Any of these motifs were confirmed to be part of an active promoter. Yet, since 50% of the identified promoters were found to deviate from these motifs, considerably more than half of all mitochondrial promoters in Arabidopsis may be expected to show divergent sequences. This underlines the necessity to experimentally define mitochondrial transcription initiation sites and the limited possibility of predicting these sites, based on conserved promoter motifs.

Defined promoter elements such as particular core sequences or the TATATA motif appear to be distributed randomly between different promoters. This precludes Arabidopsis mitochondrial promoters from being classified into distinct groups that could be related to different mitochondrial RNA polymerases, based on merely a comparison of promoter sequences. In Arabidopsis, the nucleus-encoded phage-type RNA polymerase RpoTmp is targeted not only to mitochondria but also to plastids (Hedtke, et al., 2000). This enzyme could thus be expected to recognize promoters of similar structure in both organelles. Many of the mitochondrial promoters characterized in the present study contain a CRTA core, which resembles the YRTA motif displayed by a subset of plastid promoters that are most likely used by phage-type RNA polymerases (Hess and Börner, 1999; Liere, et al., 2004). Yet, the relatively small number of plastid promoters of this type studied thus far and the variability of mitochondrial promoter structures render it difficult to tentatively assign a distinct subset of mitochondrial promoters to RpoTmp. Besides, promoter recognition by this RNA polymerase may be mediated by different, as yet unidentified cofactors in plastids and in mitochondria, and may therefore depend on different promoter sequences in the two organelles.

To study the relevance of distinct sequence elements or nucleotide positions for promoter activity, transcription from mitochondrial promoters and mutagenized promoter variants might be analyzed using an *in vitro* transcription system as described previously for maize and pea mitochondrial promoters (Dombrowski, et al., 1999; Rapp, et al., 1993). Alternatively, comparative promoter analyses might be carried out in isolated organelles. Techniques for the introduction of DNA into isolated mitochondria from maize, sorghum and

cauliflower have been developed, and the transcription of introduced genes has been shown (Farre and Araya, 2001; Staudinger, et al., 2005; Staudinger and Kempken, 2003); F. Kempken, Universität Kiel, personal communication).

Frequent duplications and rearrangements of plant mitochondrial genomes during evolution (Palmer, 1990) appear to have been important mechanisms of establishing promoter sequences throughout the mtDNA that are recognized by the nucleus-encoded transcription apparatus. Extended sequences showing high similarity to the *rrn18* 5' region and comprising both P*rrn18*-69 and P*rrn18*-156 are seen upstream of the *orf275-orf149-nad5c-nad4L-orf25* and *orf153b-orf118-orf114-nad3-rps12-orf117-ccb203* gene clusters. Although it was not tested whether transcription is initiated within the duplicated *rrn18*-like regions, it is possible that they direct transcription of the two clusters. A duplication of promoter regions is most apparent for the *atp1*, *atp6*, *atp8*, *atp9* and *rrn26* genes, which possess homologous 5' sequence stretches that comprise one or two promoters (compare promoter sequences in Table 9 and promoter distribution in Figure 31).

The idea has been put forward that in plant mitochondria, different promoter types, which could possibly be activated by distinct transcriptional cofactors, might control the transcription of particular genes or groups of genes (Mackenzie and McIntosh, 1999). However, promoter type distribution in Arabidopsis mitochondria appears to be at random, in part owing to the high frequency of mtDNA recombination events, and is thus not in support of such a mode of transcriptional control. For example, no differences are apparent between promoters of protein- or RNA-coding genes, or between genes encoding ribosomal proteins or components of the respiratory complexes.

### IV.1.3 Promoter distribution

For 9 out of 12 genes, multiple transcription initiation sites were detected (Figure 31). Six genes were found to be transcribed from three or even four promoters. Interestingly, in a series of initiation sites of one gene, promoters containing the conserved nonanucleotide motifs CGTATATAA or CRTAAGAGA were usually positioned downstream of promoters with deviating or apparently no motifs. It may be speculated that transcription is initiated at additional sites even further upstream of the defined promoters, but that it has been impossible to detect these as the experimental strategy applied in this study does probably not cover regions exceeding 2 kbp. Identifying additional upstream sites would require an extensive primer walking strategy.

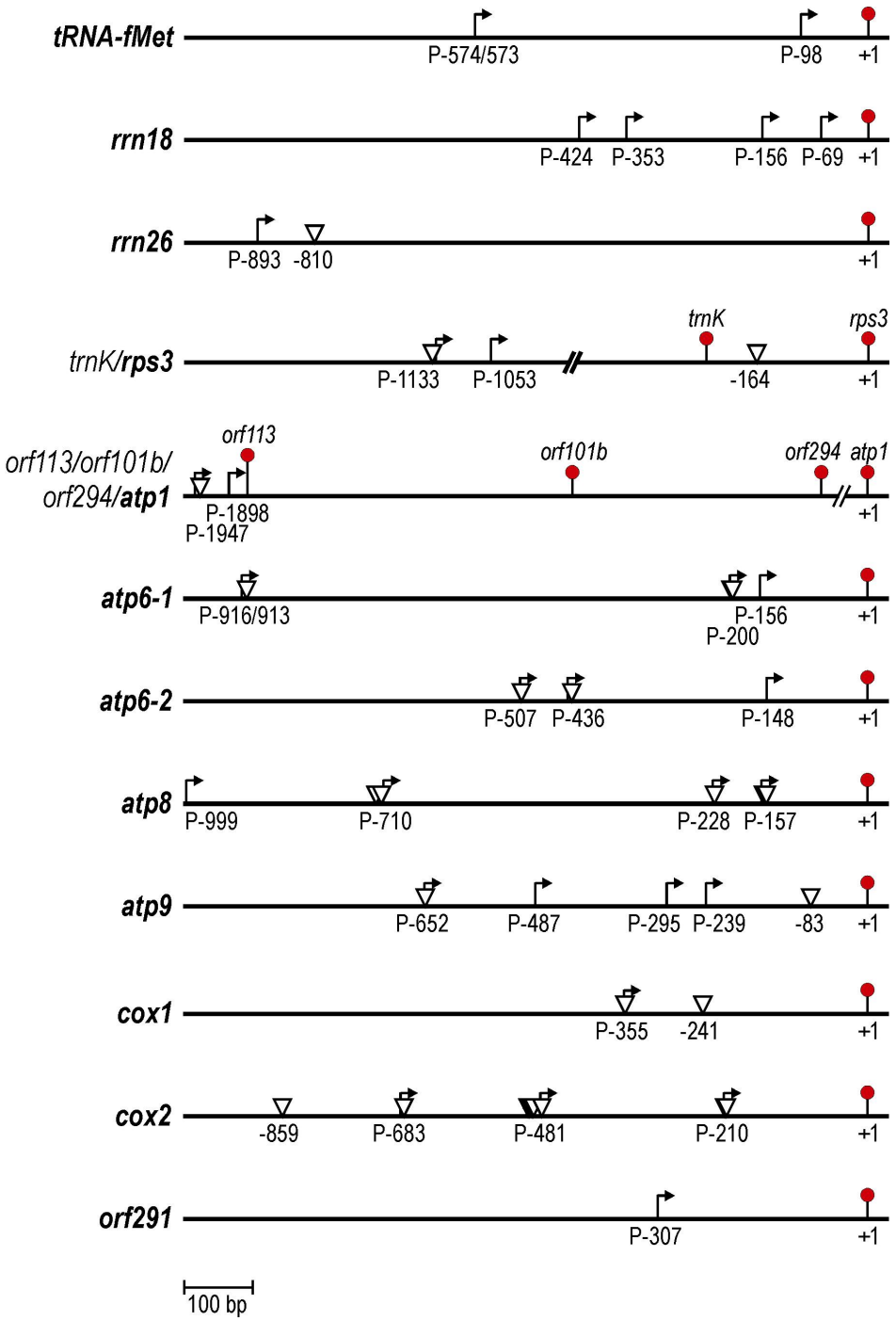



**Figure 31: Linear diagram of upstream regions of the mitochondrial genes *trnM, rrn18, rrn26, rps3, cox1, cox2, atp1, atp6-1, atp6-2, atp8, atp9* and *orf291*.** For these genes, transcript 5' ends were mapped in this study. Positions of transcriptional starts and processing sites are given as negative distances (in base pairs) from translational starts or mature rRNA or tRNA 5' ends equalling position +1, which is highlighted by a filled circle. The letter P precedes transcription initiation site positions; other symbols as in Figure 12.

While in dicots, multiple promoters have so far only been reported for the pea *cox2* gene (Kuhn and Binder, 2002), transcription initiation at multiple sites has been described for several mitochondrial genes in maize, rice and sorghum (Lupold, et al., 1999; Mulligan, et al., 1988; Nakazono, et al., 1996; Nakazono, et al., 1996; Yan and Pring, 1997). It is unlikely that

frequent promoter multiplicity, as revealed here for the first time for a dicotyledonous plant, would be restricted to Arabidopsis. 5'-RACE analyses, which easily trace even small amounts of primary transcripts, might uncover multiple promoters also in the mitochondria of other dicots.

Within the regions investigated for transcription start sites, additional promoters not displayed in Table 9 may be active from which minor primary transcripts derive. Transcripts mapping to the possible transcription initiation sites P*cox2*-231 (AGTATTATAA, see Figure 10) and P*rps3*-1149 (AGTAGATAG) did not give rise to distinct bands in the 5'-RACE, but were revealed as by-products of cloning and extensive sequencing of more abundant 5'-RACE products resulting from nearby transcriptional starts. These possible promoters are not listed in Table 9 because not enough data has been accumulated to substantiate transcription initiation at these sites. The cloning of transcript 5' ends not producing signals above the limit of detection in *in vitro* capping (P*cox2*-231) or 5'-RACE experiments implies that a multitude of minor transcription start sites within mitochondrial DNA sequences may exist.

The role of mitochondrial promoter multiplicity in plant mitochondria is hitherto unknown. While variation in mitochondrial promoter usage has been described between *Zea perennis* plants possessing different alleles of a nuclear gene (Newton, et al., 1995), there has been no evidence to date that in a single plant species with a distinct nuclear background, mitochondrial genes would be differentially transcribed. The present study shows mitochondrial transcription to initiate at identical sites in leaves and in flowers of Arabidopsis plants, indicating that tissue-specific regulation of mitochondrial genes on the level of promoter selection is of only minor or no importance in Arabidopsis. This is in line with previous conclusions that in mitochondrial gene expression, regulatory mechanisms are aimed predominantly at posttranscriptional steps (Giegé, et al., 2000). Different promoters of a distinct gene may, however, vary in their level of activity between different tissues and possibly also between different developmental stages. Quantitative analyses of primary transcripts through primer extension experiments or quantitative real-time PCR would contribute to elucidating the activity and function of individual promoters controlling one gene.

It has been suggested that multiple promoters might enable regulatory mechanisms such as differential promoter usage, producing different 5'-untranslated regions which possibly influence translational yield (Lupold, et al., 1999). Such mechanisms would be contrasted by analyses that detected a similar distribution of *cox3* transcripts with different 5'-untranslated sequences in both polysomal fractions and total RNA prepared from maize mitochondria,

implying that mitochondrial ribosomes non-preferentially associate with differently initiated transcripts (Yang and Mulligan, 1993). From the present data on primary transcript 5' termini in Arabidopsis mitochondria, a point of view is favoured that considers multiple promoters the result of a relaxed promoter specificity of the mitochondrial transcription machinery (Lupold, et al., 1999). Further experiments employing a defined Arabidopsis *in vitro* transcription system will be required to clarify whether specific promoter types are recognized by different RNA polymerase complexes containing particular transcriptional cofactors or different core enzymes.

### IV.1.4 Non-stringent control of Arabidopsis mtDNA transcription

A number of transcription initiation sites were detected in the present study that give rise to antisense transcripts to known genes or drive the expression of regions lacking identified or hypothetical ORFs, and are active in both leaves and flowers of Arabidopsis. Although it is possible that non-coding or antisense transcripts have a function in Arabidopsis mitochondria, experimental data in support of such a concept are lacking to date. It may rather be that these RNAs are the result of a relaxed transcriptional control and of non-stringent sequence requirements of plant mitochondrial promoters. In line with this, Arabidopsis mitochondrial promoters are highly diverse in architecture (compare IV.1.2), and multiple promoters driving the transcription of the same gene are apparently indifferently active (compare IV.1.3). It is possible that plant mitochondria, which undergo recurrent partial duplications and structural rearrangements of their genomes (see I.2), have accumulated transcription start sites at a fairly high frequency on the mtDNA, resulting in the synthesis of both coding and non-coding transcripts. Transcription of large non-coding regions has been described earlier in maize mitochondria (Finnegan and Brown, 1990). Although the authors did not determine whether these RNA products were due to unattenuated transcription of coding sequences or resulting from independent transcription initiation events, it is conceivable that the maize mitochondrial genome is like the Arabidopsis mtDNA transcribed from a multitude of promoters not confined to upstream regions of functional ORFs. Synthesis of transcripts from intergenic regions and of antisense RNAs has lately been reported in a variety of organisms (Johnson, et al., 2005) and is evidently not restricted to the mitochondrial compartment. Hence, as opposed to an earlier suggestion by (Lupold, et al., 1999), it may not necessarily be in response to structural peculiarities of the mtDNA that promoter multiplicity is maintained in plant mitochondria.

Support for non-functional and possibly disadvantageous transcripts being made in plant mitochondria, the synthesis of which is compensated by RNA degradation mechanisms, has recently been provided by (Holec, et al., 2005) who investigated mitochondrial RNA degradation substrates. In transgenic Arabidopsis down-regulated for polynucleotide phosphorylase (PNPase), RNAs targeted by polyadenylation for degradation by PNPase were found to include RNAs transcribed from regions that lack known functional genes (Holec, et al., 2005). Among these were transcripts of chimeric ORFs created by intragenic recombination events in the mitochondrial genome (compare I.2) as well as antisense transcripts to functional genes, which unless degraded might impede mitochondrial function. Selected RNAs were demonstrated by Northern hybridization analyses to significantly accumulate in the PNPase mutant but not in wild-type plants, and were detected in the wild type only by using more sensitive PCR techniques (Holec, et al., 2005). Plants with reduced PNPase levels were previously described to exhibit developmental defects (Perrin, et al., 2004), which may in part be due to accumulating deleterious transcripts (Holec, et al., 2005).

The seemingly non-stringent transcriptional control of plant mitochondrial gene expression thus appears to be contrasted and is possibly made up for by posttranscriptional processes. This concept may be valid for other genetic systems as similar mechanisms have been inferred from a study on transcription initiation sites in *E. coli*. Promoter-like sequences located within or downstream of ORFs on the *E. coli* chromosome were shown to activate transcription but to give rise to only minor RNA products, which might be a consequence of either imperfect promoter architecture or poor transcript stability (Kawano, et al., 2005).

## IV.2 Potential mtTFB-like cofactors of phage-type RNA polymerases in Arabidopsis

### IV.2.1 A mtTFB-like protein in Arabidopsis mitochondria

Mitochondrial transcription in yeast depends on the cofactor mtTFB in addition to the phage-type RNA polymerase Rpo41 (Lisowsky and Michaelis, 1988; Masters, et al., 1987; Schinkel, et al., 1988; Winkley, et al., 1985). An RNA polymerase of the phage type has been identified as the catalytic subunit of the transcription machinery in human mitochondria (Tiranti, et al., 1997), followed by the discovery of two human genes encoding mitochondrial mtTFB-like factors that induce promoter-specific transcription *in vitro* (Falkenberg, et al., 2002; McCulloch, et al., 2002). According to secondary and tertiary structure analyses, mtTFB is a member of the group of *S*-adenosyl-L-methionine-dependent rRNA methyltransferases (Falkenberg, et al., 2002; McCulloch, et al., 2002; Schubot, et al., 2001). No such factors have so far been described in plants although plant mitochondria similarly rely on

phage-type RNA polymerases to transcribe their genomes and are likely to require transcriptional cofactors (Hess and Börner, 1999, see I.3.3). The present study aimed at the identification of sequences encoding candidate auxiliary factors of mitochondrial RNA polymerases in Arabidopsis.

BLAST searches detected three nuclear genes in this plant encoding putative rRNA dimethylases with limited similarity to yeast and human mtTFB. Among these was *PFC1*, which has been reported earlier to encode a plastidial 16S rRNA dimethylase that is homologous to the yeast nucleolar 18S rRNA dimethylase Dim1 (Tokuhisa, et al., 1998). GFP import experiments have confirmed PFC1 to possess an N-terminal transit peptide mediating protein import into plastids of isolated tobacco protoplasts (B. Kuhla, HU Berlin, personal communication). Two other mtTFB-like genes corresponding to loci At5g66360 and At2g47420 were tentatively designated *MetA* and *MetB* respectively. Analyses of subcellular targeting demonstrated the derived MetA protein to possess an N-terminal transit peptide mediating translocation into mitochondria, thus supporting a mtTFB-like localization of MetA, whereas MetB may localize to the cytoplasm or nucleus.

Recombinant MetA and MetB expressed in and purified from *E. coli* were found to bind DNA but displayed no specificity for mtDNA fragments containing mitochondrial promoters. Not only MetA but also the presumably non-mitochondrial MetB behaved identically to human mtTFB1 in DNA binding assays (compare McCulloch, et al., 2002). It thus appears that non-specific DNA binding, which has also been reported for yeast mtTFB (Riemen and Michaelis, 1993), is a general attribute of rRNA dimethylases and rRNA-dimethylase-like proteins that is not related to the function of some members of the methyltransferase group as mitochondrial transcription factors. This idea is in line with a report on a point mutation in h-mtTFB1 that eliminated DNA binding but did not affect transcriptional activation by h-mtTFB1 *in vitro* (McCulloch and Shadel, 2003). The authors accordingly proposed a model in which h-mtTFB1 or h-mtTFB2 does not itself bind to the promoter but bridges an interaction of h-mtTFA bound to distal promoter elements with the core RNA polymerase at the transcription initiation site (see I.3.3.2 and Figure 6).

### IV.2.2 Phylogenetic relationship between mtTFB and related rRNA dimethylases

In order to resolve the relationships of MetA, MetB and Pfc1 to yeast and animal mtTFBs and to characterized rRNA dimethylases, a phylogenetic analysis was performed, which also included rRNA dimethylase-like sequences available from plants other than Arabidopsis. In line with targeting analyses (see III.2.2, Figure 16, and Annex B), Arabidopsis MetB and its

supposed nuclear or cytosolic plant orthologues were calculated to be most closely related to a group of rRNA dimethylases containing the fungal nucleolar rRNA dimethylase Dim1 as well as an additional, possibly nuclear or cytoplasmic methyltransferase from humans. A group of predicted mitochondrial rRNA dimethylases of plants, which include Arabidopsis MetA, appear to have separated from the line leading to the MetB/Dim1 cluster significantly more recently than fungal mtTFB and animal mtTFB1 and mtTFB2. Consistent with the phylogenetic relation between methyltransferase-like sequences (Figure 17), it is possible that the evolutionary process leading to the recruitment of an rRNA dimethylase as auxiliary factor of the mitochondrial transcription machinery in fungi and animals took place after the separation of the common ancestor of these groups from the green algae and plant lineage (for eukaryotic phylogeny, see Burger, et al., 2003). Thus, plant mitochondrial rRNA dimethylase-like proteins need not necessarily represent transcriptional cofactors. *In vitro* transcription studies as conducted for the human mitochondrial transcription machinery (Falkenberg, et al., 2002) and analyses of transgenic Arabidopsis lines impaired in MetA function might provide the means to clarify whether Arabidopsis MetA acts as mitochondrial transcription factor.

The position of Arabidopsis and poplar Pfc1 in the phylogram (Figure 17) is similarly inconclusive with respect to a possible additional role of these proteins as cofactors of phage-type RNA polymerases in plastids. Pfc1 appears to have separated from the line leading to the MetB/Dim1 group much earlier than MetA but is not more closely related to fungal mtTFB or animal mtTFB1 and mtTFB2 than MetA. However, in the light of the extreme divergence of fungal and animal mtTFBs, evolutionary distance to these proteins does not argue against a role of Pfc1 in plastidial transcription.

Surprisingly, fungal and animal mtTFBs do not form a single cluster in the phylogram (Figure 17). It moreover appears that mtTFB1 and mtTFB2 sequences, each forming a separate cluster containing exclusively animal proteins, are less closely related to each other than any of these groups to a cluster containing fungal nucleolar rRNA dimethylases and their putative animal orthologues. Compared to mtTFB1s, the mtTFB2 group is moderately more closely related to the highly diverse fungal mtTFBs. It thus appears that in contrast to a previous suggestion (Falkenberg, et al., 2002), mtTFB1 and mtTFB2 are not the result of a gene duplication event early in metazoan evolution and instead have separated at a prior point in time. It has been reported that human mtTFB1 is an order of magnitude less active as transcription factor than human mtTFB2 (Falkenberg, et al., 2002); the former is able to methylate rRNA at a conserved methylation site (Seidel-Rogol, et al., 2003). These observations motivated the proposition that despite the similarity of both proteins to rRNA

dimethylases and their ability to act as transcription factor *in vitro*, the methyltransferase and cofactor roles could be distributed between mtTFB1 and mtTFB2 in human mitochondria (McCulloch and Shadel, 2003). Support for this idea of biochemically non-redundant functions of mtTFB1 and mtTFB2 is gained from two studies on the roles of mtTFB1 and mtTFB2 in *Drosophila*. While cells with an RNAi-mediated reduction in mtTFB2 levels displayed defects in mitochondrial transcription and replication that could not be compensated by mtTFB1 (Matsushima, et al., 2004), RNAi knock-down of mtTFB1 reduced mitochondrial protein synthesis but had no effect on the abundance of specific mitochondrial transcripts (Matsushima, et al., 2005). The proposed roles of animal mtTFB2 as mitochondrial transcription factor and of mtTFB1 as translational modulator (Matsushima, et al., 2005) are in accordance with the phylogenetic analysis presented here.

It has been suggested earlier that the methyltransferase properties of mtTFB and its ability to bind RNA may be important for transcription factor activity (Schubot, et al., 2001). However, it has been demonstrated for human mtTFB1 that the *in vitro* function of the protein as transcription factor is independent of its rRNA dimethylase activity (McCulloch and Shadel, 2003). This emphasizes that the evolutionary process turning an rRNA dimethylase into a cofactor of a phage-type RNA polymerase may have been promoted by structural features of the methyltransferase but may not have depended on methyltransferase activity.

Considering the poor amino acid sequence conservation among fungal and animal mtTFBs (see III.2.1 and Annex A), the possibility must be considered that additional mtTFB-like sequences exist in plant genomes that encode organellar rRNA dimethylases or/and transcription factors, but which the BLAST searches conducted here have failed to detect.

### IV.2.3 Implications of the function of homologous RNA polymerases in mitochondria and plastids

The homology of nucleus-encoded plastidial and mitochondrial RNA polymerases in plants and the presence of yet another phage-type RNA polymerase in both mitochondria and plastids in dicotyledonous plants (Chang, et al., 1999; Hedtke, et al., 1997; Hedtke, et al., 2000; Hedtke, et al., 2002; Ikeda and Gray, 1999) raise the question whether the two organelles harbour similar transcriptional cofactors interacting with these core enzymes. As yet, the only characterized gene encoding a plastidial protein with similarity to known mitochondrial transcription factors in eukaryotes is the mtTFB-like rRNA dimethylase gene *PFC1*.

A transgenic Arabidopsis line carrying a T-DNA insert in *PFC1* has been reported to lack 16S rRNA methylation in the plastid at two adenosines established as conserved modification sites of the ribosome small subunit rRNA (Lafontaine, et al., 1994; Tokuhisa, et al., 1998). The characterization of this mutant had been carried out before the resolution of the three-dimensional structure of yeast mtTFB and the identification of the gene encoding human mtTFB1 uncovered these organellar transcription factors to be related to rRNA dimethylases (McCulloch, et al., 2002; Schubot, et al., 2001). Hence, the *pfc1* mutant line was not examined for possible alterations in transcripts synthesized in the plastid. Pfc1-deficient plants were of a phenotype identical to wild-type individuals at 22°C but exhibited chilling-induced chlorosis at 5°C due to impaired chloroplast development at low temperature and returned to normal chloroplast biogenesis in meristematic cells upon transfer to 22°C (Tokuhisa, et al., 1998). Pfc1 was thus concluded to be essential for chloroplast development at low temperature and to possibly compensate for a chilling-sensitive step in ribosome biogenesis. The protein may thus have a role in plastids that is equivalent to the suggested *in vivo* function of *Drosophila* mtTFB1 in mitochondria (see I.3.3.1 and IV.2.2). Further studies will be required to examine if in addition to being an rRNA dimethylase, Pfc1 is a cofactor of phage-type RNA polymerases in plastids, and if the described low-temperature defects in *pfc1* are at least partially due to imperfect transcription in the photosynthetic organelle. Analyses of transcripts in *pfc1* plastids are under way (M. Swiatecka, HU Berlin, personal communication).

No sequence encoding a putative plastidial Pfc1-like protein could be retrieved from the fully sequenced rice genome or from any other monocot species, indicating that 16S rRNA maturation and RpoT-driven transcription in monocot plastids might be independent of a methyltransferase-like protein. Alternatively, Pfc1-like sequences of monocots may have escaped the BLAST searches conducted here (compare IV.2.2).

## IV.3 Transcriptional roles of the phage-type RNA polymerases RpoTm and RpoTmp in Arabidopsis mitochondria

The RNA polymerases and auxiliary factors involved in mtDNA transcription in yeast and mammals have been identified and subjected to thorough analyses of their contribution to the transcription process (Falkenberg, et al., 2002; Gaspari, et al., 2004; Matsunaga and Jaehning, 2004; for reviews summarizing earlier studies, see Hess and Börner, 1999; Shadel and Clayton, 1993; Tracy and Stern, 1995). Reconstitution of mitochondrial transcription *in vitro* from individual recombinant components allowed defining the minimal composition of RNA

polymerase complexes capable of mitochondrial promoter recognition and transcription initiation (Falkenberg, et al., 2002; Gaspari, et al., 2004; Matsunaga and Jaehning, 2004; Matsunaga, et al., 2004). The activity of the transcription apparatus present in plant mitochondria, on the other hand, has so far mostly been characterized by analyses of transcripts synthesized *in vivo* in the mitochondrion (Hess and Börner, 1999), and references therein), or through detailed studies of *in vitro* RNA synthesis by transcriptionally active mitochondrial extracts (Binder, et al., 1995; Hanic-Joyce and Gray, 1991; Rapp and Stern, 1992). Although a biochemical dissection of such extracts has been anticipated to lead to the identification of individual components of the transcription machinery, as was attained earlier for the yeast mitochondrial RNA polymerase and its transcriptional cofactor sc-mtTFB (Greenleaf, et al., 1986; Lisowsky and Michaelis, 1988; Schinkel, et al., 1987; Winkley, et al., 1985), direct evidence is as yet lacking for specific components to be involved in mitochondrial transcription in plants. Attempts to isolate the core RNA polymerase and accessory factors have been hampered presumably mostly by the low abundance of these components in the mitochondrion and by their seemingly loose association in RNA polymerase holoenzymes.

The identification of mitochondrial promoters (this study; (Kühn, et al., 2005) and of genes encoding RNA polymerases (Hedtke, et al., 1997; Hedtke, et al., 2000) and a potential mtTFB-like cofactor (this study) of the mitochondrial transcription machinery in Arabidopsis prompted efforts to reconstitute a defined homologous *in vitro* transcription system from individual recombinant components.

### IV.3.1 Development of an Arabidopsis *in vitro* transcription system

*In vitro* transcription studies examined the abilities of RpoTm and RpoTmp to transcribe DNA from mitochondrial promoters located on supercoiled or linearized plasmid templates, and moreover tested the mitochondrial mtTFB-like protein MetA for its potential to modulate the transcriptional performances of RpoTm and RpoTmp. Transcription of neither supercoiled nor linear DNA templates by RpoTm or RpoTmp appeared to be affected by the presence of MetA in the *in vitro* assay. This contrasts the functioning of yeast mtTFB which stimulates promoter-specific transcription of both linear and supercoiled DNA (Matsunaga and Jaehning, 2004; Winkley, et al., 1985). A role of MetA that would be equivalent to that of yeast or mammalian mtTFB in mitochondrial RNA synthesis is thus not supported by the *in vitro* transcription experiments described here (see IV.4). While RpoTm recognized a variety of promoters on supercoiled DNA, identical promoters did not yield run-off RNA products if

provided on linearized templates (Table 12 and Figure 23, Figure 25, Figure 28 and Figure 30). Likewise, RpoTmp produced no specific RNAs from linear DNA. The synthesis of discrete transcripts from supercoiled plasmids by RpoTmp was confined to the promoters P*atp6-1*-916/913, P*atp6-2*-436 and P*atp6-2*-507 (Figure 28 and Figure 30). Table 12 lists promoter sequences that were included in the study and indicates their *in vitro* utilization.

*In vitro* transcription studies made use of the two bacterial ρ-independent terminator sequences *hisa* and *thra* (Barnes and Tuley, 1983; Gardner, 1982) in order to obtain RNA products of defined lengths resulting from promoter-specific initiation followed by termination at *hisa* or *thra* on circular templates (see Figure 23, Figure 25, and Figure 27 for template design). ρ-independent terminators resemble Class I terminators of the T7 phage in that they encode a G/C-rich sequence stretch followed by a run of U residues, with the former having the potential to form a stable stem loop (Farnham and Platt, 1981; Macdonald, et al., 1994; Macdonald, et al., 1993; von Hippel, 1998). A number of bacterial termination signals were shown accordingly to cause the T7 RNA polymerase to terminate (Christiansen, 1988; Jeng, et al., 1990; Jeng, et al., 1992; Steen, et al., 1986). The bacterial terminators *hisa* and *thra* seem to similarly induce termination of transcription by the T7-like enzymes RpoTm and RpoTmp *in vitro* although termination *in vivo* does apparently not involve terminator sequences in mitochondria (see I.3.2.3).

Previously described *in vitro* transcription studies routinely made use of primer extension analysis in order to characterize *in vitro* transcription product 5' ends (Binder, et al., 1995; Liere and Maliga, 1999). However, occasional discrepancies have been observed between *in vitro* transcription and primer extension signals (Binder, et al., 1995). Therefore, precise 5' end mapping was performed here on *in vitro*-synthesized RNAs employing the 5'-RACE technique used for the detection of *in vivo* transcription initiation sites in Arabidopsis mitochondria (see III.1.1). The method proved a valid tool for determining transcript 5' termini of *in vitro* transcription products.

**Table 12: Promoter recognition by RpoTm and RpoTmp in vitro.** The preference of mitochondrial promoters over random initiation sites by RpoTm and RpoTmp is indicated (+). Underlined positions mark *in vivo* start sites of transcription; upperecase nucleotides are initiation sites determined *in vitro*. Promoter cores are written bold and the frequent promoter motif TATATA(A) is highlighted.

| Gene | Promoter | Sequence | RpoTm | RpoTmp |
|---|---|---|---|---|
| *rrn18* | P*rrn18*-156 | tagaataata**cgta**tatAatcagaa | + | - |
| *atp6-1* | P*atp6-1*-200 | gccaataata**cgta**tatAagaagag | + | - |
| *atp9* | P*atp9*-295 | ctggtgctct**cgta**tatAagagaag | + | - |
| *cox2* | P*cox2*-210 | atgttggttt**cgta**tatAagaagac | + | - |
| *tRNA-fMet* | P*trnM*-98 | tttgaaatat**cgta**agaGaagaagg | + | - |
| *rrn26* | P*rrn26*-893 | ctatcaattt**cata**agaGaagaaag | + | - |
| *atp9* | P*atp9*-239 | ctatcaattt**cata**agagaagacga | - | - |
| *atp6-1* | P*atp6-1*-156 | ctatcaatct**cata**agagaagaaat | - | - |
| *atp8* | P*atp8*-157 | ctatcaatct**cata**agaGaaGaaat | (+)[a] | - |
| *rrn18* | P*rrn18*-69 | agtggaattg**aata**agagaagaaag | - | - |
| *atp8* | P*atp8*-228/226 | cataccataa**cata**tatAgaatcga | + | - |
| *atp6-1* | P*atp6-1*-916/913 | agccctttat**atta**tatAatAaagc | + | + |
| *cox2* | P*cox2*-481 | atgaatattc**atta**gatAataGatt | + | - |
| *rps3* | P*rps3*-1133 | tagaaaaaatt**atta**gtaatacgta | - | - |
| *rps3* | P*rps3*-1053 | ttttttatttt**ggta**ggtaacatcgc | - | - |
| *atp9* | P*atp9*-487 | atgtcttatt**ggta**tgtGatacaag | + | - |
| *atp9* | P*atp9*-652 | agaagattga**agta**aggagcaggtt | - | - |
| *cox2* | P**cox2*-345[b] | tctgtactgt**agta**ataGaagagtc | + | - |
| *atp6-2* | P*atp6-2*-436 | tcttgaatta**agta**tataGaaaaga | + | + |
| *atp6-2* | P*atp6-2*-507 | gataaatta**agta**tagtaatAagaa | + | + |
| *tRNA-fMet* | P*trnM*-574/573 | ctaatttatataaaaaagaccggga | - | - |

[a] Stimulation of RpoTm-driven transcription by P*atp8*-157 was evident only from 5'-RACE performed on *in vitro*-synthesized transcripts (see III.4.3).
[b] Precise mapping of initiating nucleotides was not done for discrete transcripts synthesized by RpoTmp (see III.4.4).
[c] Initiation at P**cox2*-345 was not observed *in vivo* and likely is an *in vitro* artefact (compare III.1.1 and Table 9).
[d] Results of experiments examining transcription from P*atp9*-295 and P*atp9*-239 were essentially identical to those obtained for the similar promoters P*atp6-1*-200 and P*atp6-1*-156 and are not shown in the results section.

### IV.3.2 Intrinsic promoter specificity of RpoTm

Of the promoters tested, all sequences displaying a TATATA element were recognized by RpoTm *in vitro*. Similarly, all promoters with a CGTA core stimulated transcription initiation, although P*trnM*-98, which is the only tested promoter possessing a CGTA motif but not a TATATA sequence, appeared to support initiation less efficiently. Of the promoters displaying a CATAAGAGA nonanucleotide motif, only P*rrn26*-893 gave rise to discernible though not abundant transcripts. Run-off transcription and 5' end mapping of *in vitro*-synthesized RNAs together provided no evidence for transcripts initiated specifically at P*atp6-1*-156, P*atp9*-239 or the similar P*rrn18*-69, and a minor stimulation of transcription initiation by P*atp8*-157 was revealed only by the sensitive 5'-RACE technique (compare Figure 28 and Figure 29), indicating that RpoTm does not significantly prefer these *in vivo* promoters over random start points *in vitro*. Since despite the high sequence similarity at

these promoters to P*rrn26*-893, transcription is detectably initiated only at P*rrn26*-893, it may be assumed that already minor deviations from the P*rrn26*-893 sequence impede recognition by RpoTm, or that sequences beyond the 25 nucleotides displayed in Table 12 determine promoter strength *in vitro*. Interestingly, upstream sequences extending beyond the T7 RNA polymerase promoter boundary at position -17 were recently shown to modulate promoter efficiency (Tang, et al., 2005). *In vivo* utilization of not only P*atp6-1*-156, P*atp9*-239 and P*rrn18*-69 but also P*rrn26*-893 might depend on the function of as yet unidentified transcriptional cofactors and/or extended sequence elements.

Earlier *in vitro* transcription studies using transcriptionally active extracts prepared from pea mitochondria described the faithful initiation at several promoters essentially conforming to the CRTAAGAGA consensus motif but rarely at promoters deviating from this sequence (Binder, et al., 1995; Kuhn and Binder, 2002). In the present study, *in vitro* utilization by RpoTm of promoters not possessing a CRTA sequence element but an ATTA or RGTA core was heterogeneous. While transcripts initiated at P*atp6-1*-916/913 and P*atp9*-487 had 5' termini identical to those of transcripts made from these promoters *in vivo*, initiation at P*cox2*-481 occurred at the correct nucleotide but in addition at a nearby downstream site not utilized *in vivo*. Transcript 5' ends generated from P*atp6-2*-436 and P*atp6-2*-507 mapped to positions closely downstream of the *in vivo* initiating nucleotides. This may be either due to an altered or diverse *in vitro* activity of RpoTm as compared to the *in vivo* initiation event, or caused by intrinsic problems of the 5'-RACE procedure at the ligation step. None of the *rps3* promoters gave rise to specific transcripts; neither did P*atp9*-652 or the A/T-rich P*trnM*-574/573 lacking apparent promoter sequence elements (compare IV.1.2.). The failure of both RpoTm and RpoTmp to initiate transcription at certain promoters may be due to the absence of factors activating transcription *in vivo* in the Arabidopsis *in vitro* transcription system. Such factors need not necessarily differ between dissimilar promoter sequences.

Minor transcripts that had not been detected *in vivo* were initiated by RpoTm *in vitro* at several positions of the *atp9* upstream region. Besides, one major discrete but non-specific transcript was made from the *cox2* upstream sequence. Promoter function of an mtDNA sequence *in vitro* but not *in vivo* has been observed previously in *in vitro* studies employing maize mitochondrial extracts as a source of transcription activity (Lupold, et al., 1999). Of several primary transcript 5' ends generated *in vitro* from two different maize mtDNA regions harbouring the *cox2* sequence, all except one were confirmed to also occur *in vivo*. The authors argued that rapid processing of this transcript *in vivo* might account for the discrepancy between *in vitro* transcription results and transcript 5' ends seen *in vivo*, but did

not exclude the possibility that the sequence does not serve as a promoter in the mitochondrion.

Earlier reports have put forward that not only RNA polymerase binding to promoter (and promoter-like) sequences but also escape from these sequences into elongation may be limiting to mitochondrial promoter function and transcription initiation *in vivo* (Lupold, et al., 1999; Matsunaga and Jaehning, 2004). According to a recent study of the yeast mitochondrial transcription machinery, sc-mtTFB may participate in modulating this transition. Yeast Rpo41 was shown to possess the intrinsic ability to correctly initiate transcription at mitochondrial promoters on supercoiled or pre-melted DNA templates in the absence of the accessory factor sc-mtTFB, whereas addition of the cofactor appeared to increase abortive relative to productive transcription from pre-melted templates (Matsunaga and Jaehning, 2004). Specific interaction with promoter sequence elements has moreover been attributed to the human and mouse mtTFA and mitochondrial RNA polymerase polypeptides rather than mtTFB (Gaspari, et al., 2004; compare I.3.3.2 and Figure 6). The methyltransferase-like transcriptional cofactors of animal and yeast mitochondria, which are obligatory for specific transcription initiation on linear DNA templates (Falkenberg, et al., 2002; McCulloch, et al., 2002; Winkley, et al., 1985; Xu and Clayton, 1992) owing to their contribution to DNA melting (Matsunaga and Jaehning, 2004), and which have previously been discussed as specificity factors involved in promoter recognition (Jang and Jaehning, 1991; Xu and Clayton, 1992), may thus alternatively be regarded as modulators of open promoter complex stability and escape into productive transcription from mitochondrial promoters. Correspondingly acting auxiliary factors may be part of plant mitochondrial transcription machineries, and their absence in the Arabidopsis *in vitro* transcription system may result in erratic initiation events.

In line with observations of an intrinsic promoter specificity of the yeast mitochondrial RNA polymerase Rpo41 is the accurate transcription initiation by RpoTm at diverse promoters of the Arabidopsis mtDNA, which is described here. RpoTm differs in its transcriptional performance from complex plant mitochondrial extracts in that it does not specifically initiate transcription at promoters located on linear DNA templates. This implies the participation of (an) auxiliary factor(s) in alterations of the DNA structure or in template melting and open promoter complex formation in *in vitro* transcription experiments using mitochondrial extracts (Binder, et al., 1995; Hanic-Joyce and Gray, 1991; Rapp and Stern, 1992), which according to the data presented here is obviated by a supercoiled DNA conformation that facilitates opening of the double helix.

The present study establishes the mitochondrial phage-type RNA polymerase encoded by the nuclear *RpoTm* gene to recognize mitochondrial promoters of diverse architecture *in vitro*. It thus provides the first direct linkage in the plant kingdom between an RNA polymerase activity initiating transcription at mitochondrial promoters and a nuclear gene encoding a mitochondrial phage-type RNA polymerase. *In vitro* transcription studies strongly suggest that at least for the diverse promoters shown here to stimulate initiation by RpoTm, promoter sequence specificity is conferred by the RpoTm core enzyme and does not require auxiliary factors. This questions previous ideas of different transcription factors being involved in transcription initiation at different promoter types controlling distinct genes or groups of genes (Mackenzie and McIntosh, 1999; compare also IV.1.2). Rather, general RNA polymerase cofactors and DNA-binding proteins may be involved in transcription in Arabidopsis mitochondria.

A comprehensive study directed at identifying mitochondrial DNA-binding proteins and transcription factors in Arabidopsis mitochondria will combine the computational prediction and preliminary biochemical characterization of mitochondrial DNA-binding proteins encoded by the Arabidopsis genome with a proteomics approach analyzing polypeptides attached to mitochondrial nucleoids (Thirkettle-Watts and Finnegan, 2005). Candidate transcription factors and DNA structure-modifying proteins emerging from this study could be used to complement the minimal Arabidopsis *in vitro* transcription system described here.

### IV.3.3 Different transcriptional properties of RpoTm and RpoTmp

*In vitro* transcription studies provided little information on the role of RpoTmp in Arabidopsis mitochondria. Unlike RpoTm, RpoTmp did not display a marked preference for mitochondrial promoters as transcription initiation sites. However, non-specific transcription of DNA templates was far more efficient by RpoTmp than by RpoTm (Kühn, 2001); this study). Different RNA synthesis rates and promoter specificities may be related to the *in vivo* tasks of RpoTm and RpoTmp although possible effects of the thioredoxin tag on RNA polymerase activity, which might differ between the two recombinant enzymes, should be considered. The necessity to study *in vitro* RNA synthesis by untagged RpoTm and RpoTmp is discussed below (IV.4).

The data presented here on RpoTm- and RpoTmp-dependent RNA synthesis *in vitro* are in favour of different and complementing roles of the two RNA polymerases in transcriptional processes in Arabidopsis mitochondria. Promoter recognition by RpoTmp was confined to P*atp6-1*-916/913, P*atp6-2*-436 and P*atp6-2*-507, which nevertheless induced transcription

initiation by RpoTm with higher efficiency. RNA synthesis from these promoters by RpoTmp was markedly exceeded by non-specific template transcription. Moreover, nine out of twelve promoters utilized by RpoTm did not activate RpoTmp *in vitro* (Table 12). In the light of the *in vitro* initiation at diverse mitochondrial promoters by RpoTm it is unlikely that the apparent lack of promoter specificity of RpoTmp is due to the absence of specificity factors in the *in vitro* transcription assay that would otherwise enable RpoTmp to efficiently recognize identical promoters. The weak preference displayed by RpoTmp for P*atp6-1*-916/913, P*atp6-2*-436 and P*atp6-2*-507 over random initiation sites may be conferred by conserved structural elements of phage-type RNA polymerases.

Mitochondrial as well as plastidial phage-type RNA polymerases have been observed - though not in the present study - to initiate transcription at the T7 promoter *in vitro* (Kühn, 2001; Lerbs-Mache, 1993). The sequence of the T7 promoter is entirely different from those of promoters recognized by plant phage-type RNA polymerases *in vivo*. In turn, T7 RNA polymerase has been suggested to initiate transcription at plastidial promoters of phage-type RNA polymerases following the observation that genes controlled by this promoter type showed enhanced transcript levels in transgenic tobacco overexpressing a plastidial T7 RNA polymerase (Magee and Kavanagh, 2002). It thus appears that the ability to preferentially initiate transcription at certain specific or non-specific sequences is a property that is conserved among phage and phage-type RNA polymerases, regardless of whether or not these enzymes require auxiliary factors for transcription initiation *in vivo*. Yeast Rpo41 has been proposed to possess a polypeptide region corresponding in structure, though not in sequence, to the T7 RNA polymerase specificity loop involved in promoter binding (Matsunaga and Jaehning, 2004; compare Figure 5). Corresponding structures might be formed by plant RpoT polypeptides and could contribute to an intrinsic ability of RpoT enzymes to preferentially initiate transcription at particular sequences *in vitro* (see Annex C for T7-based structural models illustrating a hypothetical folding of the RpoTm and RpoTmp polypeptides). The specificity loop region of organellar phage-type RNA polymerases shows limited sequence similarity to the T7 enzyme but is flanked by two highly conserved sequence domains. It is rich in positively charged and hydrophobic residues, supporting this structure to be involved in protein-nucleic acid interactions (Figure 32). With the aim of defining structural elements of RpoTm that confer promoter specificity, site-directed mutagenesis of recombinant RpoTm and RpoTmp could be carried out, of which the RpoT sequence region corresponding to the T7 RNA polymerase specificity loop would be a particular target. Such mutagenesis studies might well complement phylogenetic analyses of plant RpoT enzymes. Besides, it would be

of major interest to compare *in vitro* transcription activities of recombinant RpoTm and RpoTmp to those of enzymes that are active in mitochondrial lysates prepared from Arabidopsis in order to characterize functional differences between RpoT core and holoenzymes.

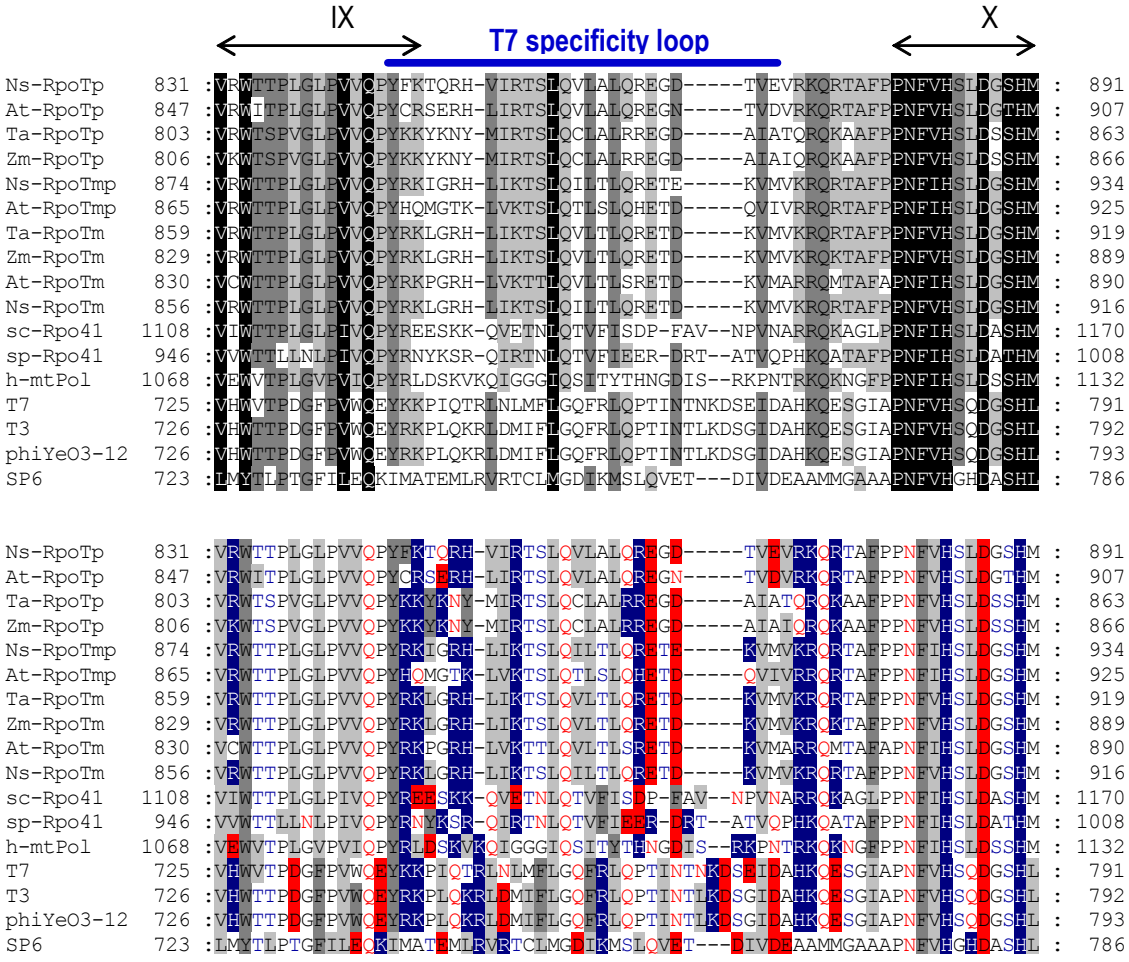



**Figure 32: Comparison of the specificity loop region of phage and phage-type RNA polymerases.** The upper aminoacid sequence alignment illustrates the level of sequence similarity between RNA polymerases of the T7 phage type in the specificity loop region (blue bar) and adjacent sequences. Shaded positions are conserved in 60 % (light grey), 80 % (grey) or 100 % (black) of aligned sequences. The lower alignment shows identical sequence regions, with amino acids shaded according to their physicochemical properties (hydrophobic, shaded light gray; positively charged, shaded blue; negatively charged, shaded red). Sequences were compared of RNA polymerases of the bacteriophages T7, T3, phiYeO3 and SP6, mitochondrial RNA polymerases from *H. sapiens* (h-mtPol), *S. cerevisiae* (sc-Rpo41), *S. pombe* (sp-Rpo41), and sequences of Arabidopsis (At-), *Nicotiana sylvestris* (Ns-), *T. aestivum* (Ta-) and *Z. mays* (Zm-) mitochondrial (RpoTm), mitochondrial/plastidial (RpoTmp) and plastidial phage-type RNA polymerases (RpoTp). The alignment was generated using the Multalin algorithm (Corpet, 1988) and refined manually. See Annex D for accession numbers.

While a role of RpoTm in transcribing genes from mitochondrial promoters can be deduced from the *in vitro* transcription studies described here, the function RpoTmp in mitochondrial transcription *in vivo* remains unresolved. It is possible that in the presence of transcriptional cofactors, RpoTmp initiates transcription of mitochondrial genes *in vivo*, perhaps at promoters not recognized by RpoTm. It may be speculated that RpoTmp functions

as a transcriptase priming mtDNA replication, and initiates transcription at as yet undefined sites associated with origins of replication. Alternatively, RpoTmp may have an as yet unreported transcriptional role that does not entail the recognition of identified mitochondrial promoters. Interestingly, preliminary data suggest both RpoTm and RpoTmp to transcribe RNA templates *in vitro* (data not shown). A similar activity has been reported for the T7-phage RNA polymerase (Konarska and Sharp, 1989; Konarska and Sharp, 1990), and has moreover been proposed to catalyze RNA replication in maize mitochondria (Formanova and Brown, 1997) and transcript 3'-end extension observed in both mitochondria and plastids (Zandueta-Criado and Bock, 2004; D. Gagliardi, IBMP CNRS, Strasbourg, France, personal communication), although the functional significance of the added sequences is unresolved.

Prior examinations of the roles of RpoTm and RpoTmp in Arabidopsis mitochondria, which predominantly addressed possible developmental or tissue-specific differences between RpoTm and RpoTmp functions, did not allow defining roles of the two RNA polymerases in mitochondrial transcription (Baba, et al., 2004; Emanuel, et al., 2005; Emanuel, et al., 2005). The recent characterization of a transgenic Arabidopsis line carrying a T-DNA insert in the last intron of the *RpoTmp* gene and apparently possessing no functional RpoTmp supported a role of the enzyme in early plastid but not in mitochondrial gene expression (Baba, et al., 2004). Data collected for both wild-type and mutant plants on plastid and mitochondrial mRNA levels at different developmental stages and under different light regimes were interpreted with difficulty, as the mutant displayed altered *RpoTm* and *RpoTp* transcript levels compared to the wild type. Phenotypic effects were therefore not directly attributable to the loss of RpoTmp function. The authors put the short-root phenotype of mutant plants down to not only the lack of RpoTmp but also reduced *RpoTm* mRNA accumulation in roots, but did not detect significant effects of the T-DNA insertion on the accumulation of mitochondrial transcripts. Based predominantly on the observation that in the mutant, the induction of several plastid genes in dark-grown mutant seedlings upon illumination was delayed, they proposed RpoTmp to be the key RNA polymerase transcribing organellar genes during early seedling development and favoured a role of both RpoTm and RpoTp at later developmental stages. However, a model in which RpoTm and RpoTmp act at different stages of development is not supported by a study showing essentially identical expression patterns of *RpoTm* and *RpoTmp* (Emanuel, et al., 2005). Baba et al. (2004) moreover suggested that RpoTm could substitute for RpoTmp in the mitochondrion, while RpoTp might partially

compensate for the loss of RpoTmp in the plastid. *In vitro* transcription studies of RpoTm and RpoTmp are compatible with this idea.

In contrast to dicots, monocots possess only one mitochondrial phage-type RNA polymerase (see I.3.3.1). The available literature is bare of data hinting at a dicot-specific aspect of mitochondrial transcription that would necessitate an additional phage-type RNA polymerase in dicot mitochondria. Thus, it is not unlikely that functions of the monocot RpoTm are in dicots distributed between RpoTm and RpoTmp, although no data exist to substantiate this proposition. Studies of mitochondrial promoter utilization and transcription in transgenic Arabidopsis plants that are reduced in RpoTm or RpoTmp protein levels or entirely deficient in RpoTm or RpoTmp function could contribute to defining distinct roles of these enzymes in dicot mitochondria.

A recent investigation of the gene encoding the human mitochondrial RNA polymerase moreover directs the search for target genes of phage-type RNA polymerases to the nuclear genome. Remarkably, this gene has been discovered to additionally give rise to an N-terminally truncated polypeptide constituting a nuclear phage-type RNA polymerase (Kravchenko, et al., 2005). This novel enzyme appears to be responsible for the transcription of a large number of nuclear protein-coding genes and initiates transcription at sequences that resemble neither the human mitochondrial promoters LSP and HSP nor promoters recognized by the nuclear RNA polymerase II (Kravchenko, et al., 2005). One could speculate that the recruitment of phage-type RNA polymerases to transcribing organellar and nuclear genes may have occurred as one evolutionary process.

## IV.4 Does a mtTFB homologue function in mitochondrial transcription in plants?

The *in vitro* transcription experiments described here strongly suggest the participation of auxiliary factors in mitochondrial RNA synthesis in Arabidopsis. Although the above described differences in template utilization between transcriptionally active plant mitochondrial extracts and RpoTm are reminiscent of the deviating template requirements observed between the yeast sc-mtTFB-Rpo41 heterodimer and isolated Rpo41 (Matsunaga and Jaehning, 2004), it is unclear whether a mtTFB homologue is present in the extracts which contributes to DNA melting or modulates transcription initiation *in vitro* and *in vivo*. The preliminary characterization of MetA, which apparently is the only mitochondrial mtTFB-like protein in Arabidopsis, through *in vitro* transcription experiments suggests that the role of MetA may not be equivalent to that of yeast and animal mtTFB (see IV.3.1). This is in line with phylogenetic analyses placing MetA and other putative mitochondrial mtTFB-

like rRNA dimethylases of plants in one group with the yeast nuclear Dim1 methyltransferase but not with experimentally confirmed mitochondrial transcription factors such as fungal or mammalian mtTFBs (see Figure 17 and IV.2.2). The data presented here do however not unequivocally contradict MetA to act as transcriptional cofactor *in vivo*. It is possible that an interaction of recombinant MetA and RpoTm in the *in vitro* transcription assay, which according to analyses of yeast and animal mtTFB is critical for mtTFB function (Cliften, et al., 2000; Falkenberg, et al., 2002), was obstructed by the thioredoxin-hexahistidine tag attached to RpoTm N-terminus. Likewise, the N-terminal hexahistidine tag of MetA may have impeded MetA function, although similar tags attached to recombinant mammalian mtTFB were not reported to be disadvantageous (Gaspari, et al., 2004). Two-hybrid analyses examining a possible association of MetA with RpoTm or RpoTmp were not in favour of heterodimer formation (data not shown). Yet, fusion partners attached to MetA and RpoT proteins in these studies may have hindered interaction in the two-hybrid assay. On the other hand, a two-hybrid assay has been successfully used to monitor sc-mtTFB-Rpo41 interaction in yeast (Cliften, et al., 2000). *In vitro* transcription studies using untagged RpoT enzymes, which could be prepared from sources described in chapter III.3, will be required to more adequately examine the ability of MetA to interact with RpoTm or RpoTmp and to modulate RpoTm- or RpoTmp-dependent transcription. An Arabidopsis mutant line carrying a T-DNA insertion in the MetA coding sequence has recently become available from the GABI-Kat mutant collection (http://www.gabi-kat.de/). Mitochondrial transcript analyses of this mutant and of transgenic plants expressing interfering RNAs directed at MetA transcript levels may help elucidate a possible role of MetA in mitochondrial transcription. Alternatively, MetA may merely be functioning as rRNA dimethylase in Arabidopsis mitochondria, whereas cofactors mediating transcription initiation in plant mitochondria might not be of the rRNA dimethylase type. Presumed 18S rRNA methylation by MetA is presently investigated by S. Okada and A. Brennicke (Universität Ulm). The possibility that in addition to MetA, a mtTFB homologue functioning as transcription factor exists in Arabidopsis mitochondria which escaped the BLAST searches conducted here, has been discussed in section IV.2.2.

### IV.5 Transcription initiation in Arabidopsis mitochondria

How could RpoTm initiate transcription at promoters in Arabidopsis mitochondria? In the presence of the native mitochondrial nucleoid, which is composed of not only DNA but also DNA-binding proteins of as yet unknown identity, the DNA may have a conformation that allows the RNA polymerase to initiate transcription at mitochondrial promoters without the

aid of additional auxiliary factors (Figure 33, model A). However, *in vitro* studies using plant mitochondrial extracts as a source of transcription activity imply that (a) soluble cofactor(s) may be involved in promoter-specific transcription initiation (Binder, et al., 1995; Fey, et al., 1999; Tracy and Stern, 1995). As in yeast mitochondria, these cofactors which may be functionally equivalent but not similar in their structure to yeast mtTFB, could associate with RpoTm prior to promoter binding (Figure 33, model B). The cofactor-independent *in vitro* specificity of RpoTm for diverse promoter sequences (see Table 12) suggests that in order to initiate transcription in Arabidopsis mitochondria, RpoTm associates with general auxiliary factors rather than promoter sequence-specific cofactors. Similar factors might enable RpoTmp to initiate transcription at mitochondrial promoters. Considering that both plant mitochondrial extracts and recombinant RpoTm recognize promoters of different architecture but not all *in vivo* promoter sequences *in vitro* (Binder, et al., 1995); this study), it may be argued that different promoters could require distinct modes of transcription initiation. Such mechanisms are however unlikely to represent a major means of mitochondrial gene regulation since studies of mitochondrial promoter utilization *in vivo* suggest a non-stringent control of transcription initiation (see IV.1). Thus, mitochondrial genes in Arabidopsis may not be individually regulated at the transcriptional level.

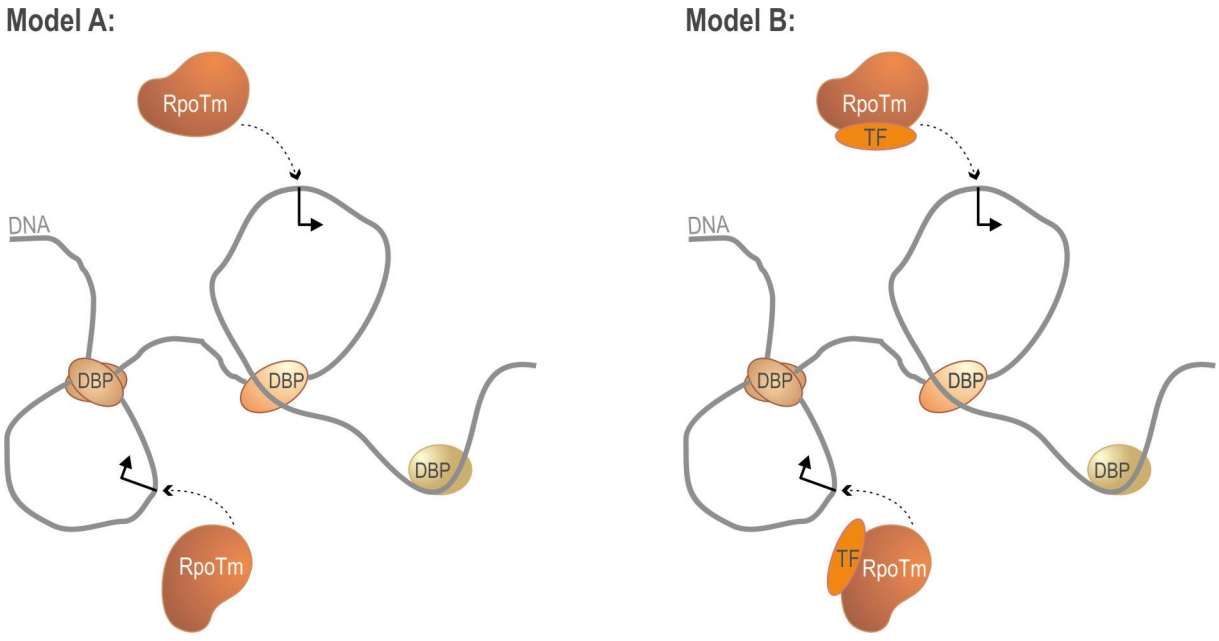



**Figure 33: Transcription initiation in Arabidopsis mitochondria.** Association of RpoTm with promoter sequences (bent arrows) and transcription initiation *in vivo* may depend on hypothetical DNA-binding proteins (DBP) determining the in vivo nucleoid structure and/or on as yet unidentified general transcriptional cofactors binding the core RNA polymerase in a 1:1 stochiometry (TF, model B).

# V REFERENCES

Abdelnoor, R. V.; Yule, R.; Elo, A.; Christensen, A. C.; Meyer-Gauen, G. and Mackenzie, S. A. (2003): Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS, Proc Natl Acad Sci U S A 100 [10], pp. 5968-5973.

Adams, K. L.; Daley, D. O.; Qiu, Y. L.; Whelan, J. and Palmer, J. D. (2000): Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants, Nature 408 [6810], pp. 354-357.

Adams, K. L.; Qiu, Y. L.; Stoutemyer, M. and Palmer, J. D. (2002): Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution, Proc Natl Acad Sci U S A 99 [15], pp. 9905-9912.

Adams, K. L.; Song, K.; Roessler, P. G.; Nugent, J. M.; Doyle, J. L.; Doyle, J. J. and Palmer, J. D. (1999): Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes, Proc Natl Acad Sci U S A 96 [24], pp. 13863-13868.

Alam, T. I.; Kanki, T.; Muta, T.; Ukaji, K.; Abe, Y.; Nakayama, H.; Takio, K.; Hamasaki, N. and Kang, D. (2003): Human mitochondrial DNA is packaged with TFAM, Nucleic Acids Res 31 [6], pp. 1640-1645.

Anderson, S.; Bankier, A. T.; Barrell, B. G.; de Bruijn, M. H.; Coulson, A. R.; Drouin, J.; Eperon, I. C.; Nierlich, D. P.; Roe, B. A.; Sanger, F.; Schreier, P. H.; Smith, A. J.; Staden, R. and Young, I. G. (1981): Sequence and organization of the human mitochondrial genome, Nature 290 [5806], pp. 457-465.

Andre, C.; Levy, A. and Walbot, V. (1992): Small repeated sequences and the structure of plant mitochondrial genomes, Trends Genet 8 [4], pp. 128-132.

Antoshechkin, I. and Bogenhagen, D. F. (1995): Distinct roles for two purified factors in transcription of *Xenopus* mitochondrial DNA, Mol Cell Biol 15 [12], pp. 7032-7042.

Antoshechkin, I.; Bogenhagen, D. F. and Mastrangelo, I. A. (1997): The HMG-box mitochondrial transcription factor xl-mtTFA binds DNA as a tetramer to activate bidirectional transcription, EMBO J 16 [11], pp. 3198-3206.

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H. and Altuvia, S. (2001): Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*, Curr Biol 11 [12], pp. 941-950.

Baba, K.; Schmidt, J.; Espinosa-Ruiz, A.; Villarejo, A.; Shiina, T.; Gardestrom, P.; Sane, A. P. and Bhalerao, R. P. (2004): Organellar gene transcription and early seedling development are affected in the rpoT;2 mutant of Arabidopsis, Plant J 38 [1], pp. 38-48.

Backert, S. and Börner, T. (2000): Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.), Curr Genet 37 [5], pp. 304-314.

Backert, S.; Nielsen, B. L. and Börner, T. (1997): The mystery of the rings: structure and replication of mitochondrial genomes from higher plants, Trends Plant Sci 2 [12], pp. 477-483.

Bannai, H.; Tamada, Y.; Maruyama, O.; Nakai, K. and Miyano, S. (2002): Extensive feature detection of N-terminal protein sorting signals, Bioinformatics 18 [2], pp. 298-305.

Barnes, W. M. and Tuley, E. (1983): DNA sequence changes of mutations in the histidine operon control region that decrease attenuation, J Mol Biol 165 [3], pp. 443-459.

Bates, P. A.; Kelley, L. A.; MacCallum, R. M. and Sternberg, M. J. (2001): Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM, Proteins Suppl 5, pp. 39-46.

Bates, P. A. and Sternberg, M. J. (1999): Model building by comparison at CASP3: using expert knowledge and computer automation, Proteins Suppl 3, pp. 47-54.

Beardslee, T. A.; Roy-Chowdhury, S.; Jaiswal, P.; Buhot, L.; Lerbs-Mache, S.; Stern, D. B. and Allison, L. A. (2002): A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts, Plant J 31 [2], pp. 199-209.

Bendich, A. J. (1993): Reaching for the ring: the study of mitochondrial genome structure, Curr Genet 24 [4], pp. 279-290.

Bensing, B. A.; Meyer, B. J. and Dunny, G. M. (1996): Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of *Enterococcus faecalis*, Proc Natl Acad Sci U S A 93 [15], pp. 7794-7799.

Binder, S. and Brennicke, A. (2003): Gene expression in plant mitochondria: transcriptional and post-transcriptional control, Philos Trans R Soc Lond B Biol Sci 358 [1429], pp. 181-188; discussion 188-189.

Binder, S.; Hatzack, F. and Brennicke, A. (1995): A novel pea mitochondrial in vitro transcription system recognizes homologous and heterologous mRNA and tRNA promoters, J Biol Chem 270 [38], pp. 22182-22189.

Binder, S.; Marchfelder, A. and Brennicke, A. (1996): Regulation of gene expression in plant mitochondria, Plant Mol Biol 32 [1-2], pp. 303-314.

Binder, S.; Thalheim, C. and Brennicke, A. (1994): Transcription of potato mitochondrial 26S rRNA is initiated at its mature 5' end, Curr Genet 26 [5-6], pp. 519-523.

Bligny, M.; Courtois, F.; Thaminy, S.; Chang, C. C.; Lagrange, T.; Baruah-Wolff, J.; Stern, D. and Lerbs-Mache, S. (2000): Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerases, EMBO J 19 [8], pp. 1851-1860.

Bogenhagen, D. F. (1996): Interaction of mtTFB and mtRNA polymerase at core promoters for transcription of *Xenopus laevis* mtDNA, J Biol Chem 271 [20], pp. 12036-12041.

Bogenhagen, D. F. and Insdorf, N. F. (1988): Purification of *Xenopus laevis* mitochondrial RNA polymerase and identification of a dissociable factor required for specific transcription, Mol Cell Biol 8 [7], pp. 2910-296.

Boore, J. L. (1999): Animal mitochondrial genomes, Nucleic Acids Res 27 [8], pp. 1767-1780.

Bradford, M. M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal Biochem 72, pp. 248-254.

Brennicke, A.; Zabaleta, E.; Dombrowski, S.; Hoffmann, M. and Binder, S. (1999): Transcription signals of mitochondrial and nuclear genes for mitochondrial proteins in dicot plants, J Hered 90 [3], pp. 345-350.

Brown, G. G.; Auchincloss, A. H.; Covello, P. S.; Gray, M. W.; Menassa, R. and Singh, M. (1991): Characterization of transcription initiation sites on the soybean mitochondrial genome allows identification of a transcription-associated sequence motif, Mol Gen Genet 228 [3], pp. 345-355.

Bullerwell, C. E. and Gray, M. W. (2004): Evolution of the mitochondrial genome: protist connections to animals, fungi and plants, Curr Opin Microbiol 7 [5], pp. 528-534.

Burger, G.; Gray, M. W. and Lang, B. F. (2003): Mitochondrial genomes: anything goes, Trends Genet 19 [12], pp. 709-716.

Cahoon, A. B.; Harris, F. M. and Stern, D. B. (2004): Analysis of developing maize plastids reveals two mRNA stability classes correlating with RNA polymerase type, EMBO Rep 5 [8], pp. 801-806.

Cannon, G. C.; Ward, L. N.; Case, C. I. and Heinhorst, S. (1999): The 68 kDa DNA compacting nucleoid protein from soybean chloroplasts inhibits DNA synthesis in vitro, Plant Mol Biol 39 [4], pp. 835-845.

Caoile, A. G. and Stern, D. B. (1997): A conserved core element is functionally important for maize mitochondrial promoter activity in vitro, Nucleic Acids Res 25 [20], pp. 4055-4060.

Cermakian, N.; Ikeda, T. M.; Cedergren, R. and Gray, M. W. (1996): Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage, Nucleic Acids Res 24 [4], pp. 648-654.

Cermakian, N.; Ikeda, T. M.; Miramontes, P.; Lang, B. F.; Gray, M. W. and Cedergren, R. (1997): On the evolution of the single-subunit RNA polymerases, J Mol Evol 45 [6], pp. 671-681.

Chang, C. C.; Sheen, J.; Bligny, M.; Niwa, Y.; Lerbs-Mache, S. and Stern, D. B. (1999): Functional analysis of two maize cDNAs encoding T7-like RNA polymerases, Plant Cell 11 [5], pp. 911-926.

Cheetham, G. M.; Jeruzalmi, D. and Steitz, T. A. (1999): Structural basis for initiation of transcription from an RNA polymerase-promoter complex, Nature 399 [6731], pp. 80-83.

Cheetham, G. M. and Steitz, T. A. (1999): Structure of a transcribing T7 RNA polymerase initiation complex, Science 286 [5448], pp. 2305-2309.

Christiansen, J. (1988): The 9S RNA precursor of *Escherichia coli* 5S RNA has three structural domains: implications for processing, Nucleic Acids Res 16 [15], pp. 7457-7476.

Claros, M. G. and Vincens, P. (1996): Computational method to predict mitochondrially imported proteins and their targeting sequences, Eur J Biochem 241 [3], pp. 779-786.

Cliften, P. F.; Jang, S. H. and Jaehning, J. A. (2000): Identifying a core RNA polymerase surface critical for interactions with a sigma-like specificity factor, Mol Cell Biol 20 [18], pp. 7013-7023.

Cliften, P. F.; Park, J. Y.; Davis, B. P.; Jang, S. H. and Jaehning, J. A. (1997): Identification of three regions essential for interaction between a sigma-like factor and core RNA polymerase, Genes Dev 11 [21], pp. 2897-2909.

Contreras-Moreira, B. and Bates, P. A. (2002): Domain fishing: a first step in protein comparative modelling, Bioinformatics 18 [8], pp. 1141-1142.

Corpet, F. (1988): Multiple sequence alignment with hierarchical clustering, Nucleic Acids Res 16 [22], pp. 10881-10890.

Covello, P. S. and Gray, M. W. (1991): Sequence analysis of wheat mitochondrial transcripts capped in vitro: definitive identification of transcription initiation sites, Curr Genet 20 [3], pp. 245-251.

Crooks, G. E.; Hon, G.; Chandonia, J. M. and Brenner, S. E. (2004): WebLogo: a sequence logo generator, Genome Res 14 [6], pp. 1188-1190.

Dai, H.; Lo, Y. S.; Litvinchuk, A.; Wang, Y. T.; Jane, W. N.; Hsiao, L. J. and Chiang, K. S. (2005): Structural and functional characterizations of mung bean mitochondrial nucleoids, Nucleic Acids Res 33 [15], pp. 4725-4739.

Dairaghi, D. J.; Shadel, G. S. and Clayton, D. A. (1995): Addition of a 29 residue carboxyl-terminal tail converts a simple HMG box-containing protein into a transcriptional activator, J Mol Biol 249 [1], pp. 11-28.

Dairaghi, D. J.; Shadel, G. S. and Clayton, D. A. (1995): Human mitochondrial transcription factor A and promoter spacing integrity are required for transcription initiation, Biochim Biophys Acta 1271 [1], pp. 127-134.

Däschner, K.; Couee, I. and Binder, S. (2001): The mitochondrial isovaleryl-coenzyme a dehydrogenase of arabidopsis oxidizes intermediates of leucine and valine catabolism, Plant Physiol 126 [2], pp. 601-612.

Delarue, M.; Poch, O.; Tordo, N.; Moras, D. and Argos, P. (1990): An attempt to unify the structure of polymerases, Protein Eng 3 [6], pp. 461-467.

Diffley, J. F. and Stillman, B. (1991): A close relative of the nuclear, chromosomal high-mobility group protein HMG1 in yeast mitochondria, Proc Natl Acad Sci U S A 88 [17], pp. 7864-7868.

Diffley, J. F. and Stillman, B. (1992): DNA binding properties of an HMG1-related protein from yeast mitochondria, J Biol Chem 267 [5], pp. 3368-3374.

Dombrowski, S.; Brennicke, A. and Binder, S. (1997): 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators, EMBO J 16 [16], pp. 5069-5076.

Dombrowski, S.; Hoffmann, M.; Guha, C. and Binder, S. (1999): Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria, J Biol Chem 274 [15], pp. 10094-10099.

Dombrowski, S.; Hoffmann, M.; Kuhn, J.; Brennicke, A. and Binder, S. (1998): On mitochondrial promoters in *Arabidopsis thaliana* and other flowering plants, Möller, I. M.; Glaser, E. and Glimelius, K., Plant mitochondria pp. 165-170, Backhuys Publishers, The Netherlands.

Dower, W. J.; Miller, J. F. and Ragsdale, C. W. (1988): High efficiency transformation of E. coli by high voltage electroporation, Nucleic Acids Res 16 [13], pp. 6127-6145.

Duchene, A. M. and Marechal-Drouard, L. (2001): The chloroplast-derived trnW and trnM-e genes are not expressed in Arabidopsis mitochondria, Biochem Biophys Res Commun 285 [5], pp. 1213-1216.

Edqvist, J. and Bergman, P. (2002): Nuclear identity specifies transcriptional initiation in plant mitochondria, Plant Mol Biol 49 [1], pp. 59-68.

Elo, A.; Lyznik, A.; Gonzalez, D. O.; Kachman, S. D. and Mackenzie, S. A. (2003): Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the Arabidopsis genome, Plant Cell 15 [7], pp. 1619-1631.

Emanuel, C.; von Groll, U; Müller, M.; Börner, T. and Weihe, A. (2005): Development- and tissue-specific expression of the *RpoT* gene family of Arabidopsis encoding mitochondrial and plastid RNA polymerases, submitted.

Emanuel, C.; von Groll, U.; Muller, M.; Börner, T. and Weihe, A. (2005): Development- and tissue-specific expression of the RpoT gene family of Arabidopsis encoding mitochondrial and plastid RNA polymerases, Planta, pp. 1-12.

Emanuel, C.; Weihe, A.; Graner, A.; Hess, W. R. and Börner, T. (2004): Chloroplast development affects expression of phage-type RNA polymerases in barley leaves, Plant J 38 [3], pp. 460-472.

Emanuelsson, O.; Nielsen, H.; Brunak, S. and von Heijne, G. (2000): Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, J Mol Biol 300 [4], pp. 1005-1016.

Falkenberg, M.; Gaspari, M.; Rantanen, A.; Trifunovic, A.; Larsson, N. G. and Gustafsson, C. M. (2002): Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA, Nat Genet 31 [3], pp. 289-294.

Fangman, W. L. and Dujon, B. (1984): Yeast mitochondrial genomes consisting of only A.T base pairs replicate and exhibit suppressiveness, Proc Natl Acad Sci U S A 81 [22], pp. 7156-7160.

Farnham, PJ and Platt, T (1981): Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro, Nucl Acids Res 9 [3], pp. 563-577.

Farre, Jean-Claude and Araya, Alejandro (2001): Gene expression in isolated plant mitochondria: high fidelity of transcription, splicing and editing of a transgene product in electroporated organelles, Nucl Acids Res 29 [12], pp. 2484-2491.

Fauron, C.; Casper, M.; Gao, Y. and Moore, B. (1995): The maize mitochondrial genome: dynamic, yet functional, Trends Genet 11 [6], pp. 228-235.

Fey, J. and Marechal-Drouard, L. (1999): Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria, Biochem Biophys Res Commun 256 [2], pp. 409-414.

Fey, J.; Vermel, M.; Grienenberger, J.; Marechal-Drouard, L. and Gualberto, J. M. (1999): Characterization of a plant mitochondrial active chromosome, FEBS Lett 458 [2], pp. 124-128.

Filee, J. and Forterre, P. (2005): Viral proteins functioning in organelles: a cryptic origin?, Trends Microbiol.

Finnegan, P. M. and Brown, G. G. (1990): Transcriptional and Post-Transcriptional Regulation of RNA Levels in Maize Mitochondria, Plant Cell 2 [1], pp. 71-83.

Fisher, R. P. and Clayton, D. A. (1985): A transcription factor required for promoter recognition by human mitochondrial RNA polymerase. Accurate initiation at the heavy- and light-strand promoters dissected and reconstituted in vitro, J Biol Chem 260 [20], pp. 11330-11338.

Fisher, R. P. and Clayton, D. A. (1988): Purification and characterization of human mitochondrial transcription factor 1, Mol Cell Biol 8 [8], pp. 3496-3509.

Fisher, R. P.; Lisowsky, T.; Parisi, M. A. and Clayton, D. A. (1992): DNA wrapping and bending by a mitochondrial high mobility group-like transcriptional activator protein, J Biol Chem 267 [5], pp. 3358-3367.

Fisher, R. P.; Topper, J. N. and Clayton, D. A. (1987): Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements, Cell 50 [2], pp. 247-258.

Formanova, N. and Brown, G. G. (1997): The maize mitochondrial plasmid RNA b is associated with protein during synthesis but is not encapsidated, Plant Mol Biol 34 [3], pp. 383-392.

Gardner, JF (1982): Initiation, pausing, and termination of transcription in the threonine operon regulatory region of *Escherichia coli*, J Biol Chem 257 [7], pp. 3896-3904.

Gaspari, M.; Falkenberg, M.; Larsson, N. G. and Gustafsson, C. M. (2004): The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells, EMBO J 23 [23], pp. 4606-4614.

Giegé, P.; Hoffmann, M.; Binder, S. and Brennicke, A. (2000): RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria, EMBO Rep 1 [2], pp. 164-170.

Giegé, P.; Sweetlove, L. J.; Cognat, V. and Leaver, C. J. (2005): Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis, Plant Cell 17 [5], pp. 1497-1512.

Giese, A.; Thalheim, C.; Brennicke, A. and Binder, S. (1996): Correlation of nonanucleotide motifs with transcript initiation of 18S rRNA genes in mitochondria of pea, potato and Arabidopsis, Mol Gen Genet 252 [4], pp. 429-436.

Giese, K.; Pagel, J. and Grosschedl, R. (1997): Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1, Proc Natl Acad Sci U S A 94 [24], pp. 12845-12850.

Gray, J. C.; Sullivan, J. A.; Wang, J. H.; Jerome, C. A. and MacLean, D. (2003): Coordination of plastid and nuclear gene expression, Philos Trans R Soc Lond B Biol Sci 358 [1429], pp. 135-44; discussion 144-5.

Gray, M. W. (1992): The endosymbiont hypothesis revisited, Int Rev Cytol 141, pp. 233-357.

Gray, M. W.; Burger, G. and Lang, B. F. (1999): Mitochondrial evolution, Science 283 [5407], pp. 1476-1481.

Gray, M. W. and Lang, B. F. (1998): Transcription in chloroplasts and mitochondria: a tale of two polymerases, Trends Microbiol 6 [1], pp. 1-3.

Greenleaf, A. L.; Kelly, J. L. and Lehman, I. R. (1986): Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome, Proc Natl Acad Sci U S A 83 [10], pp. 3391-3394.

Grosschedl, R.; Giese, K. and Pagel, J. (1994): HMG domain proteins: architectural elements in the assembly of nucleoprotein structures, Trends Genet 10 [3], pp. 94-100.

Guex, N. and Peitsch, M. C. (1997): SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 [15], pp. 2714-2723.

Hanaoka, M.; Kanamaru, K.; Fujiwara, M.; Takahashi, H. and Tanaka, K. (2005): Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis, EMBO Rep 6 [6], pp. 545-550.

Handa, H. (2003): The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*, Nucleic Acids Res 31 [20], pp. 5907-5916.

Hanic-Joyce, P. J. and Gray, M. W. (1991): Accurate transcription of a plant mitochondrial gene *in vitro*, Mol Cell Biol 11 [4], pp. 2035-2039.

Hanson, M. R. and Bentolila, S. (2004): Interactions of mitochondrial and nuclear genes that affect male gametophyte development, Plant Cell 16 Suppl, pp. S154-169.

Hatzack, F.; Dombrowski, S.; Brennicke, A. and Binder, S. (1998): Characterization of DNA-Binding Proteins from Pea Mitochondria, Plant Physiol 116 [2], pp. 519-528.

He, B.; Kukarin, A.; Temiakov, D.; Chin-Bow, S. T.; Lyakhov, D. L.; Rong, M.; Durbin, R. K. and McAllister, W. T. (1998): Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase, J Biol Chem 273 [30], pp. 18802-18811.

Hedtke, B. (1998): Isolation und Charakterisierung bakteriophagenähnlicher organellärer RNA-Polymerasen aus höheren Pflanzen, PhD Thesis, HU Berlin.

Hedtke, B.; Börner, T. and Weihe, A. (1997): Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis, Science 277 [5327], pp. 809-811.

Hedtke, B.; Börner, T. and Weihe, A. (2000): One RNA polymerase serving two genomes, EMBO Rep 1 [5], pp. 435-440.

Hedtke, B.; Legen, J.; Weihe, A.; Herrmann, R. G. and Börner, T. (2002): Six active phage-type RNA polymerase genes in *Nicotiana tabacum*, Plant J 30 [6], pp. 625-637.

Hedtke, B.; Meixner, M.; Gillandt, S.; Richter, E.; Börner, T. and Weihe, A. (1999): Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants *in vivo*, Plant J 17 [5], pp. 557-561.

Hedtke, B.; Wagner, I.; Börner, T. and Hess, W. R. (1999): Inter-organellar crosstalk in higher plants: impaired chloroplast development affects mitochondrial gene and transcript levels, Plant J 19 [6], pp. 635-643.

Heinhorst, S.; Chi-Ham, C. L.; Adamson, S. W. and Cannon, G. C. (2004): The somatic inheritance of plant organelles, Daniell, H. and Chase, C. D., Molecular biology and biotechnology of plant organelles pp. 37-92, Springer.

Herrmann, J. M. (2003): Converting bacteria to organelles: evolution of mitochondrial protein sorting, Trends Microbiol 11 [2], pp. 74-79.

Hess, W. R. and Börner, T. (1999): Organellar RNA polymerases of higher plants, Int Rev Cytol 190, pp. 1-59.

Hoffmann, M. and Binder, S. (2002): Functional importance of nucleotide identities within the pea atp9 mitochondrial promoter sequence, J Mol Biol 320 [5], pp. 943-950.

Holec, S.; Lange, H.; Kühn, K.; Alioua, M.; Börner, T. and Gagliardi, D. (2005): Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and PNPase, submitted.

Huang, J.; Struck, F.; Matzinger, D. F. and Levings, C. S., 3rd (1994): Flower-enhanced expression of a nuclear-encoded mitochondrial respiratory protein is associated with changes in mitochondrion number, Plant Cell 6 [3], pp. 439-448.

Ikeda, T. M. and Gray, M. W. (1999): Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria, Mol Cell Biol 19 [12], pp. 8113-8122.

Ikeda, T. M. and Gray, M. W. (1999): Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat, Plant Mol Biol 40 [4], pp. 567-578.

The Arabidopsis Genome Initiative (2000): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 [6814], pp. 796-815.

Jang, S. H. and Jaehning, J. A. (1991): The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors, J Biol Chem 266 [33], pp. 22671-22677.

Jeng, S. T.; Gardner, J. F. and Gumport, R. I. (1990): Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators, J Biol Chem 265 [7], pp. 3823-3830.

Jeng, S. T.; Gardner, J. F. and Gumport, R. I. (1992): Transcription termination in vitro by bacteriophage T7 RNA polymerase. The role of sequence elements within and surrounding a rho-independent transcription terminator, J Biol Chem 267 [27], pp. 19306-19312.

Jeruzalmi, D. and Steitz, T. A. (1998): Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme, EMBO J 17 [14], pp. 4101-4113.

Johnson, J. M.; Edwards, S.; Shoemaker, D. and Schadt, E. E. (2005): Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments, Trends Genet 21 [2], pp. 93-102.

Jones, D. T. (1999): GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences, J Mol Biol 287 [4], pp. 797-815.

Jones, D. T.; Taylor, W. R. and Thornton, J. M. (1992): The rapid generation of mutation data matrices from protein sequences, Comput Appl Biosci 8 [3], pp. 275-282.

Kabeya, Y.; Hashimoto, K. and Sato, N. (2002): Identification and characterization of two phage-type RNA polymerase cDNAs in the moss *Physcomitrella patens*: implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants, Plant Cell Physiol 43 [3], pp. 245-255.

Kabeya, Y. and Sato, N. (2005): Unique translation initiation at the second AUG codon determines mitochondrial localization of the phage-type RNA polymerases in the moss *Physcomitrella patens*, Plant Physiol 138 [1], pp. 369-382.

Kanamaru, K. and Tanaka, K. (2004): Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants, Biosci Biotechnol Biochem 68 [11], pp. 2215-2223.

Kanki, T.; Ohgaki, K.; Gaspari, M.; Gustafsson, C. M.; Fukuoh, A.; Sasaki, N.; Hamasaki, N. and Kang, D. (2004): Architectural role of mitochondrial transcription factor A in maintenance of human mitochondrial DNA, Mol Cell Biol 24 [22], pp. 9823-9834.

Kawano, Mitsuoki; Storz, Gisela; Rao, B. Sridhar; Rosner, Judah L. and Martin, Robert G. (2005): Detection of low-level promoter activity within open reading frame sequences of *Escherichia coli*, Nucl Acids Res 33 [19], pp. 6268-6276.

Kobayashi, Y.; Dokiya, Y.; Kumazawa, Y. and Sugita, M. (2002): Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in *Nicotiana sylvestris*, Biochem Biophys Res Commun 299 [1], pp. 57-61.

Kobayashi, Y.; Dokiya, Y. and Sugita, M. (2001): Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts, Biochem Biophys Res Commun 289 [5], pp. 1106-1113.

Kobayashi, Y.; Dokiya, Y.; Sugiura, M.; Niwa, Y. and Sugita, M. (2001): Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in *Nicotiana sylvestris*, Gene 279 [1], pp. 33-40.

Konarska, Maria M. and Sharp, Phillip A. (1989): Replication of RNA by the DNA-dependent RNA polymerase of phage T7, Cell 57 [3], pp. 423-431.

Konarska, Maria M. and Sharp, Phillip A. (1990): Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase, Cell 63 [3], pp. 609-618.

Kravchenko, J. E.; Rogozin, I. B.; Koonin, E. V. and Chumakov, P. M. (2005): Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin, Nature 436 [7051], pp. 735-739.

Kubo, T.; Nishizawa, S.; Sugawara, A.; Itchoda, N.; Estiati, A. and Mikami, T. (2000): The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA), Nucleic Acids Res 28 [13], pp. 2571-2576.

Kuhn, J. and Binder, S. (2002): RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of cox2 transcripts in pea mitochondria, Nucleic Acids Res 30 [2], pp. 439-446.

Kühn, K. (2001): T7-phagenverwandte Transkriptionssysteme in *Arabidopsis thaliana*: Funktionelle Charakterisierung rekombinanter Komponenten, Diploma thesis, HU Berlin.

Kühn, K.; Weihe, A. and Börner, T. (2005): Multiple promoters are a common feature of mitochondrial genes in Arabidopsis, Nucleic Acids Res 33 [1], pp. 337-346.

Laemmli, U. K. (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 [5259], pp. 680-685.

Lafontaine, D.; Delcour, J.; Glasser, A. L.; Desgres, J. and Vandenhaute, J. (1994): The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast, J Mol Biol 241 [3], pp. 492-497.

Leon, P.; Arroyo, A. and Mackenzie, S. (1998): Nuclear Control of Plastid and Mitochondrial Development in Higher Plants, Annu Rev Plant Physiol Plant Mol Biol 49, pp. 453-480.

Lerbs-Mache, S. (1993): The 110-kDa polypeptide of spinach plastid DNA-dependent RNA polymerase: single-subunit enzyme or catalytic core of multimeric enzyme complexes?, Proc Natl Acad Sci U S A 90 [12], pp. 5509-5513.

Li, X. Q.; Zhang, M. and Brown, G. G. (1996): Cell-Specific Expression of Mitochondrial Transcripts in Maize Seedlings, Plant Cell 8 [11], pp. 1961-1975.

Liere, K. and Maliga, P. (1999): In vitro characterization of the tobacco *rpoB* promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters, EMBO J 18 [1], pp. 249-257.

Liere, Karsten; Kaden, Daniela; Maliga, Pal and Börner, Thomas (2004): Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type, Nucl Acids Res 32 [3], pp. 1159-1165.

Linke, Bettina and Börner, Thomas (2005): Mitochondrial effects on flower and pollen development, Mitochondrion In Press.

Lisowsky, T. and Michaelis, G. (1988): A nuclear gene essential for mitochondrial replication suppresses a defect of mitochondrial transcription in *Saccharomyces cerevisiae*, Mol Gen Genet 214 [2], pp. 218-223.

Lisowsky, T.; Wilkens, D.; Stein, T.; Hedtke, B.; Borner, T. and Weihe, A. (2002): The C-terminal region of mitochondrial single-subunit RNA polymerases contains species-specific determinants for maintenance of intact mitochondrial genomes, Mol Biol Cell 13 [7], pp. 2245-2255.

Lizama, L.; Holuigue, L. and Jordana, X. (1994): Transcription initiation sites for the potato mitochondrial gene coding for subunit 9 of ATP synthase (*atp9*), FEBS Lett 349 [2], pp. 243-248.

Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R. and Wright, P. E. (1995): Structural basis for DNA bending by the architectural transcription factor LEF-1, Nature 376 [6543], pp. 791-795.

Lupold, D. S.; Caoile, A. G. and Stern, D. B. (1999): The maize mitochondrial cox2 gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units, J Biol Chem 274 [6], pp. 3897-3903.

Lurin, C.; Andres, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyere, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; Lecharny, A.; Le Ret, M.; Martin-Magniette, M. L.; Mireau, H.; Peeters, N.; Renou, J. P.; Szurek, B.; Taconnat, L. and Small, I. (2004): Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, Plant Cell 16 [8], pp. 2089-2103.

Lyakhov, D. L.; He, B.; Zhang, X.; Studier, F. W.; Dunn, J. J. and McAllister, W. T. (1998): Pausing and termination by bacteriophage T7 RNA polymerase, J Mol Biol 280 [2], pp. 201-213.

Macdonald, L. E.; Durbin, R. K.; Dunn, J. J. and McAllister, W. T. (1994): Characterization of two types of termination signal for bacteriophage T7 RNA polymerase, J Mol Biol 238 [2], pp. 145-158.

Macdonald, L. E.; Zhou, Y. and McAllister, W. T. (1993): Termination and slippage by bacteriophage T7 RNA polymerase, J Mol Biol 232 [4], pp. 1030-1047.

Mackenzie, S. and McIntosh, L. (1999): Higher plant mitochondria, Plant Cell 11 [4], pp. 571-586.

Magee, A. M. and Kavanagh, T. A. (2002): Plastid genes transcribed by the nucleus-encoded plastid RNA polymerase show increased transcript accumulation in transgenic plants expressing a chloroplast-localized phage T7 RNA polymerase, J Exp Bot 53 [379], pp. 2341-2349.

Mangus, D. A.; Jang, S. H. and Jaehning, J. A. (1994): Release of the yeast mitochondrial RNA polymerase specificity factor from transcription complexes, J Biol Chem 269 [42], pp. 26568-26574.

Marchfelder, A.; Brennicke, A. and Binder, S. (1996): RNA editing is required for efficient excision of tRNA(Phe) from precursors in plant mitochondria, J Biol Chem 271 [4], pp. 1898-1903.

Margulis, L. (1970): Origin of Eukaryotic Cells, Yale Univ. Press, New Haven, CT.

Margulis, L. (1981): Symbiosis in Cell Evolution, Freeman, San Francisco.

Marienfeld, J.; Unseld, M. and Brennicke, A. (1999): The mitochondrial genome of Arabidopsis is composed of both native and immigrant information, Trends Plant Sci 4 [12], pp. 495-502.

Martin, W. and Herrmann, R. G. (1998): Gene transfer from organelles to the nucleus: how much, what happens, and Why?, Plant Physiol 118 [1], pp. 9-17.

Martinez-Zapater, J. M.; Gil, P.; Capel, J. and Somerville, C. R. (1992): Mutations at the Arabidopsis *CHM* locus promote rearrangements of the mitochondrial genome, Plant Cell 4 [8], pp. 889-899.

Masters, B. S.; Stohl, L. L. and Clayton, D. A. (1987): Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7, Cell 51 [1], pp. 89-99.

Matsunaga, M. and Jaehning, J. A. (2004): Intrinsic promoter recognition by a "core" RNA polymerase, J Biol Chem 279 [43], pp. 44239-44242.

Matsunaga, M. and Jaehning, J. A. (2004): A mutation in the yeast mitochondrial core RNA polymerase, Rpo41, confers defects in both specificity factor interaction and promoter utilization, J Biol Chem 279 [3], pp. 2012-2019.

Matsunaga, M.; Jang, S. H. and Jaehning, J. A. (2004): Expression and purification of wild type and mutant forms of the yeast mitochondrial core RNA polymerase, Rpo41, Protein Expr Purif 35 [1], pp. 126-130.

Matsushima, Y.; Adan, C.; Garesse, R. and Kaguni, L. S. (2005): Drosophila mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells, J Biol Chem 280 [17], pp. 16815-16820.

Matsushima, Y.; Garesse, R. and Kaguni, L. S. (2004): Drosophila mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in schneider cells, J Biol Chem 279 [26], pp. 26900-26905.

McAllister, W. T. and Raskin, C. A. (1993): The phage RNA polymerases are related to DNA polymerases and reverse transcriptases, Mol Microbiol 10 [1], pp. 1-6.

McCulloch, V.; Seidel-Rogol, B. L. and Shadel, G. S. (2002): A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine, Mol Cell Biol 22 [4], pp. 1116-1125.

McCulloch, V. and Shadel, G. S. (2003): Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity, Mol Cell Biol 23 [16], pp. 5816-5824.

McGuffin, L. J.; Bryson, K. and Jones, D. T. (2000): The PSIPRED protein structure prediction server, Bioinformatics 16 [4], pp. 404-405.

McGuffin, L. J. and Jones, D. T. (2003): Improvement of the GenTHREADER method for genomic fold recognition, Bioinformatics 19 [7], pp. 874-881.

Millar, A. H.; Day, A. D. and Whelan, J. (2004): Mitochondrial biogenesis and function in Arabidopsis, Somerville, C. R. and Meyerowitz, E. M., The Arabidopsis Book, American Society of Plant Biologists, Rockville, MD.

Miyagi, T.; Kapoor, S.; Sugita, M. and Sugiura, M. (1998): Transcript analysis of the tobacco plastid operon rps2/atpI/H/F/A reveals the existence of a non-consensus type II (NCII) promoter upstream of the atpI coding sequence, Mol Gen Genet 257 [3], pp. 299-307.

Morgan, M. K. and Ow, D. W. (1995): Polyethylene glycol-mediated transfection of tobacco leaf mesophyll protoplasts: An experiment in the study of cre-lox recombination, Maliga, P.; Klessig, D. F.; Cashmore, A. R.; Gruissem, W. and Varner, J. E., Methods in Plant Molecular Biology: A Laboratory Course Manual p. 1–17, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Mulligan, R. M.; Lau, G. T. and Walbot, V. (1988): Numerous transcription initiation sites exist for the maize mitochondrial genes for subunit 9 of the ATP synthase and subunit 3 of cytochrome oxidase, Proc Natl Acad Sci U S A 85 [21], pp. 7998-8002.

Mulligan, R. M.; Leon, P. and Walbot, V. (1991): Transcriptional and posttranscriptional regulation of maize mitochondrial gene expression, Mol Cell Biol 11 [1], pp. 533-543.

Murray, M. G. and Thompson, W. F. (1980): Rapid isolation of high molecular weight plant DNA, Nucleic Acids Res 8 [19], pp. 4321-4325.

Nakazono, M.; Ishikawa, M.; Yoshida, K. T.; Tsutsumi, N. and Hirai, A. (1996): Multiple initiation sites for transcription of a gene for subunit 1 of F1-ATPase (*atp1*) in rice mitochondria, Curr Genet 29 [5], pp. 417-422.

Nakazono, M.; Nishiwaki, S.; Tsutsumi, N. and Hirai, A. (1996): A chloroplast-derived sequence is utilized as a source of promoter sequences for the gene for subunit 9 of NADH dehydrogenase (*nad9*) in rice mitochondria, Mol Gen Genet 252 [4], pp. 371-378.

Newton, K. J.; Winberg, B.; Yamato, K.; Lupold, S. and Stern, D. B. (1995): Evidence for a novel mitochondrial promoter preceding the *cox2* gene of perennial teosintes, EMBO J 14 [3], pp. 585-593.

Nielsen, H.; Engelbrecht, J.; Brunak, S. and von Heijne, G. (1997): Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng 10 [1], pp. 1-6.

Notsu, Y.; Masood, S.; Nishikawa, T.; Kubo, N.; Akiduki, G.; Nakazono, M.; Hirai, A. and Kadowaki, K. (2002): The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants, Mol Genet Genomics 268 [4], pp. 434-445.

Oldenburg, D. J. and Bendich, A. J. (1996): Size and Structure of Replicating Mitochondrial DNA in Cultured Tobacco Cells, Plant Cell 8 [3], pp. 447-461.

Oldenburg, D. J. and Bendich, A. J. (2001): Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein, J Mol Biol 310 [3], pp. 549-562.

Palmer, J. D. (1990): Contrasting modes and tempos of genome evolution in land plant organelles, Trends Genet 6 [4], pp. 115-120.

Palmer, J. D.; Adams, K. L.; Cho, Y.; Parkinson, C. L.; Qiu, Y. L. and Song, K. (2000): Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates, Proc Natl Acad Sci U S A 97 [13], pp. 6960-6966.

Parisi, M. A. and Clayton, D. A. (1991): Similarity of human mitochondrial transcription factor 1 to high mobility group proteins, Science 252 [5008], pp. 965-969.

Parisi, M. A.; Xu, B. and Clayton, D. A. (1993): A human mitochondrial transcriptional activator can functionally replace a yeast mitochondrial HMG-box protein both in vivo and in vitro, Mol Cell Biol 13 [3], pp. 1951-1961.

Peeters, N. M.; Chapron, A.; Giritch, A.; Grandjean, O.; Lancelin, D.; Lhomme, T.; Vivrel, A. and Small, I. (2000): Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin, J Mol Evol 50 [5], pp. 413-423.

Peitsch, M. C. (1995): Protein modeling by E-mail, Bio/Technology 13, pp. 658-660.

Perrin, R.; Meyer, E. H.; Zaepfel, M.; Kim, Y. J.; Mache, R.; Grienenberger, J. M.; Gualberto, J. M. and Gagliardi, D. (2004): Two exoribonucleases act sequentially to process mature 3'-ends of atp9 mRNAs in Arabidopsis mitochondria, J Biol Chem 279 [24], pp. 25440-25446.

Pfannschmidt, T. and Liere, K. (2005): Redox regulation and modification of proteins controlling chloroplast gene expression, Antioxid Redox Signal 7 [5-6], pp. 607-618.

Phinney, B. S. and Thelen, J. J. (2005): Proteomic characterization of a triton-insoluble fraction from chloroplasts defines a novel group of proteins associated with macromolecular structures, J Proteome Res 4 [2], pp. 497-506.

Poyton, R. O. and McEwen, J. E. (1996): Crosstalk between nuclear and mitochondrial genomes, Annu Rev Biochem 65, pp. 563-607.

Rantanen, A.; Gaspari, M.; Falkenberg, M.; Gustafsson, C. M. and Larsson, N. G. (2003): Characterization of the mouse genes for mitochondrial transcription factors B1 and B2, Mamm Genome 14 [1], pp. 1-6.

Rapp, W. D.; Lupold, D. S.; Mack, S. and Stern, D. B. (1993): Architecture of the maize mitochondrial *atp1* promoter as determined by linker-scanning and point mutagenesis, Mol Cell Biol 13 [12], pp. 7232-7238.

Rapp, W. D. and Stern, D. B. (1992): A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial *atp1* gene, EMBO J 11 [3], pp. 1065-1073.

Richter, U.; Kiessling, J.; Hedtke, B.; Decker, E.; Reski, R.; Borner, T. and Weihe, A. (2002): Two RpoT genes of *Physcomitrella patens* encode phage-type RNA polymerases with dual targeting to mitochondria and plastids, Gene 290 [1-2], pp. 95-105.

Riemen, G. and Michaelis, G. (1993): A point mutation in the core subunit gene of yeast mitochondrial RNA polymerase is suppressed by a high level of specificity factor MTF1, Mol Gen Genet 237 [1-2], pp. 49-57.

Rodermel, S. (2001): Pathways of plastid-to-nucleus signaling, Trends Plant Sci 6 [10], pp. 471-478.

Rong, M.; He, B.; McAllister, W. T. and Durbin, R. K. (1998): Promoter specificity determinants of T7 RNA polymerase, Proc Natl Acad Sci U S A 95 [2], pp. 515-519.

Ronquist, Fredrik and Huelsenbeck, John P. (2003): MrBayes 3: Bayesian phylogenetic inference under mixed models, Bioinformatics 19 [12], pp. 1572-1574.

Sambrook, J. and Russell, D. (2001): Molecular Cloning: A Laboratory Manual, 3. ed., Cold Spring Harbor Laboratory Press.

Saraste, M. (1999): Oxidative phosphorylation at the fin de siecle, Science 283 [5407], pp. 1488-1493.

Sato, N.; Nakayama, M. and Hase, T. (2001): The 70-kDa major DNA-compacting protein of the chloroplast nucleoid is sulfite reductase, FEBS Lett 487 [3], pp. 347-350.

Sato, N.; Terasawa, K.; Miyajima, K. and Kabeya, Y. (2003): Organization, developmental dynamics, and evolution of plastid nucleoids, Int Rev Cytol 232, pp. 217-262.

Schinkel, A. H.; Groot Koerkamp, M. J. and Tabak, H. F. (1988): Mitochondrial RNA polymerase of *Saccharomyces cerevisiae*: composition and mechanism of promoter recognition, Embo J 7 [10], pp. 3255-3262.

Schinkel, A. H.; Groot Koerkamp, M. J.; Teunissen, A. W. and Tabak, H. F. (1988): RNA polymerase induces DNA bending at yeast mitochondrial promoters, Nucleic Acids Res 16 [19], pp. 9147-9163.

Schinkel, A. H.; Koerkamp, M. J.; Touw, E. P. and Tabak, H. F. (1987): Specificity factor of yeast mitochondrial RNA polymerase. Purification and interaction with core RNA polymerase, J Biol Chem 262 [26], pp. 12785-12791.

Schneider, T. D. and Stephens, R. M. (1990): Sequence logos: a new way to display consensus sequences, Nucleic Acids Res 18 [20], pp. 6097-6100.

Schubot, F. D.; Chen, C. J.; Rose, J. P.; Dailey, T. A.; Dailey, H. A. and Wang, B. C. (2001): Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription, Protein Sci 10 [10], pp. 1980-1988.

Schwede, T.; Kopp, J.; Guex, N. and Peitsch, M. C. (2003): SWISS-MODEL: An automated protein homology-modeling server, Nucleic Acids Res 31 [13], pp. 3381-3385.

Seidel-Rogol, B. L.; McCulloch, V. and Shadel, G. S. (2003): Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop, Nat Genet 33 [1], pp. 23-24.

Sekine, K.; Hase, T. and Sato, N. (2002): Reversible DNA compaction by sulfite reductase regulates transcriptional activity of chloroplast nucleoids, J Biol Chem 277 [27], pp. 24399-24404.

Shadel, G. S. and Clayton, D. A. (1993): Mitochondrial transcription initiation. Variation and conservation, J Biol Chem 268 [22], pp. 16083-16086.

Shadel, G. S. and Clayton, D. A. (1995): A *Saccharomyces cerevisiae* mitochondrial transcription factor, sc-mtTFB, shares features with sigma factors but is functionally distinct, Mol Cell Biol 15 [4], pp. 2101-2108.

Shadel, G. S. and Clayton, D. A. (1997): Mitochondrial DNA maintenance in vertebrates, Annu Rev Biochem 66, pp. 409-435.

Shen, E. L. and Bogenhagen, D. F. (2001): Developmentally-regulated packaging of mitochondrial DNA by the HMG-box protein mtTFA during *Xenopus* oogenesis, Nucleic Acids Res 29 [13], pp. 2822-2828.

Small, I. D. and Peeters, N. (2000): The PPR motif - a TPR-related motif prevalent in plant organellar proteins, Trends Biochem Sci 25 [2], pp. 46-47.

Small, I.; Peeters, N.; Legeai, F. and Lurin, C. (2004): Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences, Proteomics 4 [6], pp. 1581-1590.

Smart, C. J.; Moneger, F. and Leaver, C. J. (1994): Cell-specific regulation of gene expression in mitochondria during anther development in sunflower, Plant Cell 6 [6], pp. 811-825.

Sousa, R. (1996): Structural and mechanistic relationships between nucleic acid polymerases, Trends Biochem Sci 21 [5], pp. 186-190.

Sousa, R.; Chung, Y. J.; Rose, J. P. and Wang, B. C. (1993): Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution, Nature 364 [6438], pp. 593-599.

Staudinger, M.; Bolle, N. and Kempken, F. (2005): Mitochondrial electroporation and in organello RNA editing of chimeric *atp6* transcripts, Mol Genet Genomics 273 [2], pp. 130-136.

Staudinger, M. and Kempken, F. (2003): Electroporation of isolated higher-plant mitochondria: transcripts of an introduced *cox2* gene, but not an *atp6* gene, are edited *in organello*, Mol Genet Genomics 269 [4], pp. 553-561.

Steen, R.; Dahlberg, A. E.; Lade, B. N.; Studier, F. W. and Dunn, J. J. (1986): T7 RNA polymerase directed expression of the *Escherichia coli rrnB* operon, Embo J 5 [5], pp. 1099-1103.

Steitz, T. A. (2004): The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase, Curr Opin Struct Biol 14 [1], pp. 4-9.

Sugiyama, Y.; Watase, Y.; Nagase, M.; Makita, N.; Yagura, S.; Hirai, A. and Sugiura, M. (2005): The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants, Mol Genet Genomics 272 [6], pp. 603-615.

Tahirov, T. H.; Temiakov, D.; Anikin, M.; Patlan, V.; McAllister, W. T.; Vassylyev, D. G. and Yokoyama, S. (2002): Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution, Nature 420 [6911], pp. 43-50.

Tandara, H. (2000): *In vitro*-Transkriptions- und *in vivo*-Importstudien zur Charakterisierung von Transkriptionsfaktoren pflanzlicher Organellen, Diploma thesis, HU Berlin.

Tang, G. Q.; Bandwar, R. P. and Patel, S. S. (2005): Extended upstream A-T sequence increases T7 promoter strength, J Biol Chem.

Thirkettle-Watts, D. and Finnegan, P. (2005): Identifying mitochondrial transcription factors in Arabidopsis, International Congress on Plant Mitochondrial Biology, Obernai, France.

Tiranti, V.; Savoia, A.; Forti, F.; D'Apolito, M. F.; Centra, M.; Rocchi, M. and Zeviani, M. (1997): Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database, Hum Mol Genet 6 [4], pp. 615-625.

Tokuhisa, J. G.; Vijayan, P.; Feldmann, K. A. and Browse, J. A. (1998): Chloroplast development at low temperatures requires a homolog of *DIM1*, a yeast gene encoding the 18S rRNA dimethylase, Plant Cell 10 [5], pp. 699-711.

Tracy, R. L. and Stern, D. B. (1995): Mitochondrial transcription initiation: promoter structures and RNA polymerases, Curr Genet 28 [3], pp. 205-216.

Unseld, M.; Marienfeld, J. R.; Brandt, P. and Brennicke, A. (1997): The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides, Nat Genet 15 [1], pp. 57-61.

Vogel, J.; Axmann, I. M.; Herzel, H. and Hess, W. R. (2003): Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4, Nucleic Acids Res 31 [11], pp. 2890-2899.

von Hippel, Peter H. (1998): An Integrated Model of the Transcription Complex in Elongation, Termination, and Editing, Science 281 [5377], pp. 660-665.

Wang, Y. and Shadel, G. S. (1999): Stability of the mitochondrial genome requires an amino-terminal domain of yeast mitochondrial RNA polymerase, Proc Natl Acad Sci U S A 96 [14], pp. 8046-8051.

Ward, B. L.; Anderson, R. S. and Bendich, A. J. (1981): The mitochondrial genome is large and variable in a family of plants (*cucurbitaceae*), Cell 25 [3], pp. 793-803.

Weihe, A. and Börner, T. (1999): Transcription and the architecture of promoters in chloroplasts, Trends Plant Sci 4 [5], pp. 169-170.

Weihe, A.; Hedtke, B. and Börner, T. (1997): Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album*, Nucleic Acids Res 25 [12], pp. 2319-2325.

Winkley, C. S.; Keller, M. J. and Jaehning, J. A. (1985): A multicomponent mitochondrial RNA polymerase from *Saccharomyces cerevisiae*, J Biol Chem 260 [26], pp. 14214-14223.

Woody, A. Y.; Eaton, S. S.; Osumi-Davis, P. A. and Woody, R. W. (1996): Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription, Biochemistry 35 [1], pp. 144-152.

Xu, B. and Clayton, D. A. (1992): Assignment of a yeast protein necessary for mitochondrial transcription initiation, Nucleic Acids Res 20 [5], pp. 1053-1059.

Yan, B. and Pring, D. R. (1997): Transcriptional initiation sites in sorghum mitochondrial DNA indicate conserved and variable features, Curr Genet 32 [4], pp. 287-295.

Yang, A. J. and Mulligan, R. M. (1993): Distribution of maize mitochondrial transcripts in polysomal RNA: evidence for non-selectivity in recruitment of mRNAs, Curr Genet 23 [5-6], pp. 532-536.

Yang, D.; Oyaizu, Y.; Oyaizu, H.; Olsen, G. J. and Woese, C. R. (1985): Mitochondrial origins, Proc Natl Acad Sci U S A 82 [13], pp. 4443-4447.

Yin, Y. W. and Steitz, T. A. (2002): Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase, Science 298 [5597], pp. 1387-1395.

Yin, Y. W. and Steitz, T. A. (2004): The structural mechanism of translocation and helicase activity in T7 RNA polymerase, Cell 116 [3], pp. 393-404.

Young, D. A.; Allen, R. L.; Harvey, A. J. and Lonsdale, D. M. (1998): Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced, Mol Gen Genet 260 [1], pp. 30-37.

Young, D.A. and Lonsdale, D.M. (1997): Evidence that plant mitochondrial transcription requires promoter-specific factors, Maize Genet Coop News 71, pp. 69-71.

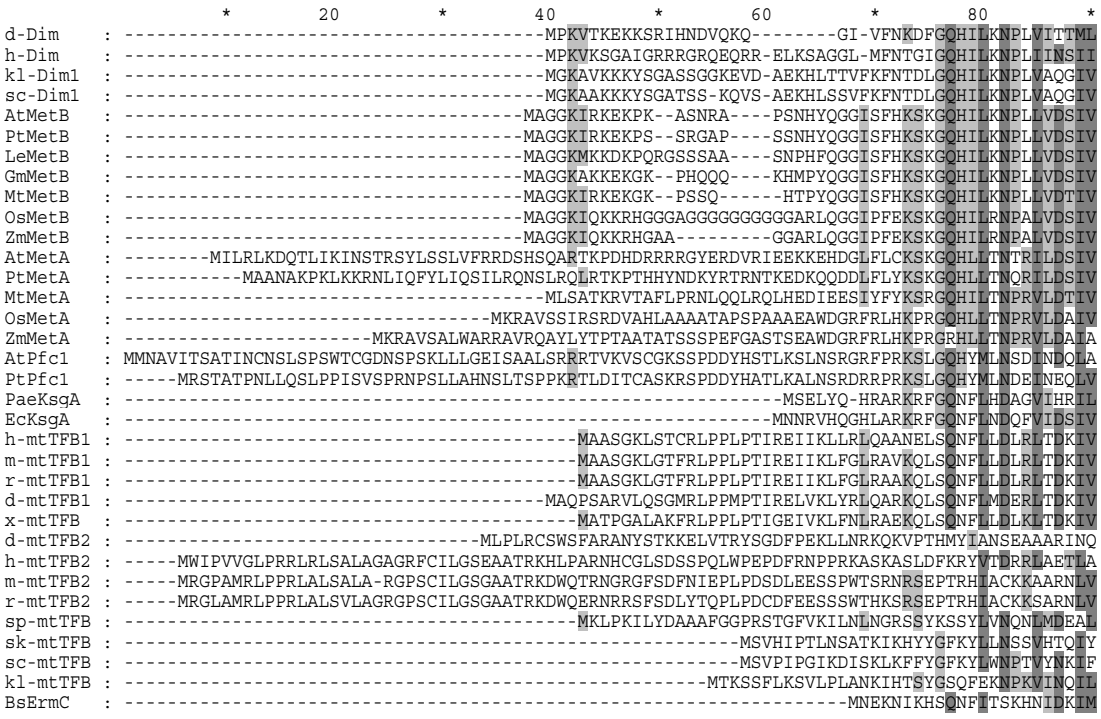
Zandueta-Criado, A. and Bock, R. (2004): Surprising features of plastid *ndhD* transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon, Nucleic Acids Res 32 [2], pp. 542-550.

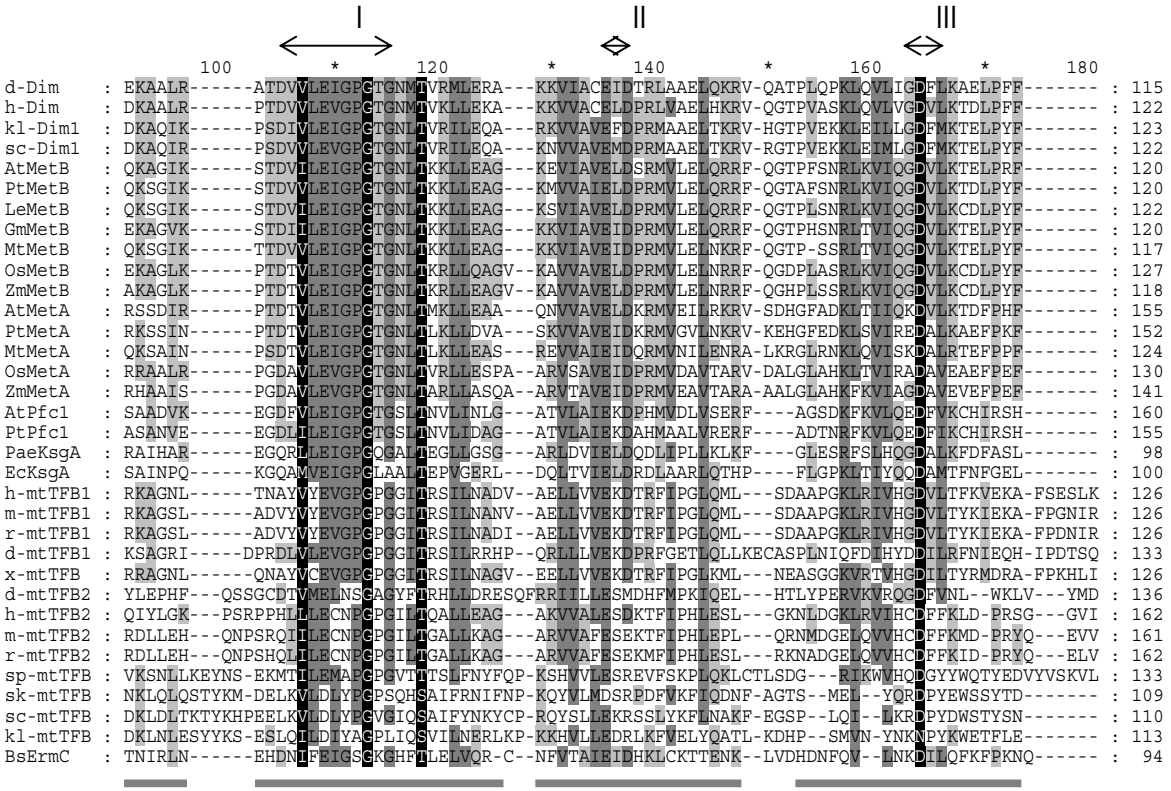
## Abbreviations

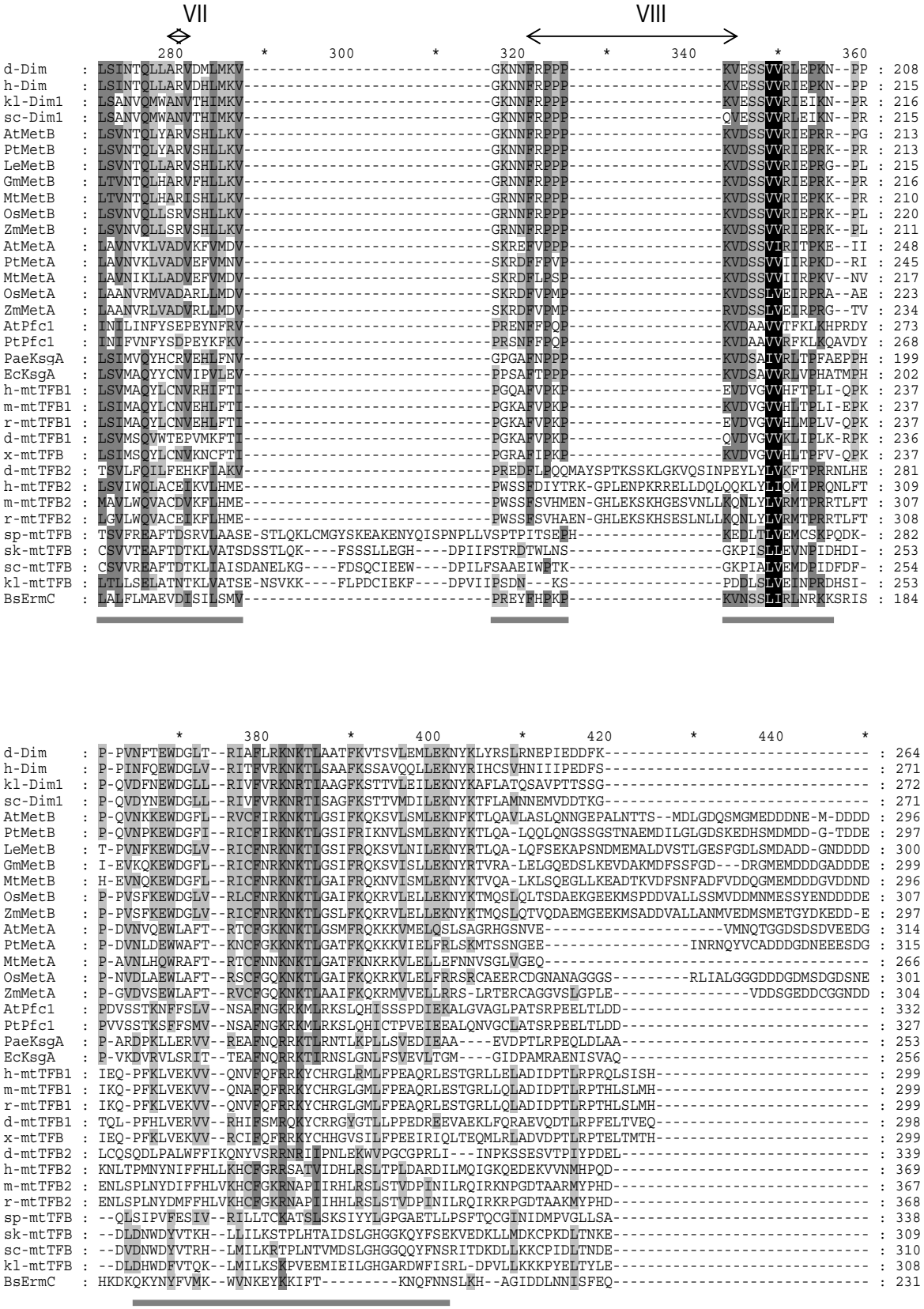
| | |
|---|---|
| aa | amino acid |
| A, C, G, T, U | nucleid acid bases (adenine, cytosine, guanine, thymin, uracil) |
| (d)ATP | (deoxy)adenosine triphosphate |
| BCIP | 5-Bromo-4-chloro-3-indolyl phosphate |
| BSA | bovine serum albumin |
| bp | base pairs |
| cDNA | complementary DNA |
| cRNA | complementary RNA |
| CTAB | cetyl trimethyl ammonium bromide |
| (d)CTP | (deoxy)cytidine triphosphate |
| DNA | deoxyribonucelic acid |
| DTT | dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| EST | expressed sequence tag |
| g | gravitational acceleration |
| GFP | green fluorescent protein |
| (d)GTP | (deoxy)guanosine triphosphate |
| h | hour |
| His | histidine |
| kDa | kilodalton |
| kbp | kilobase pairs |
| min | minute |
| MOPS | morpholinopropan-sulfonic acid |
| mRNA | messenger RNA |
| mtDNA | mitochondrial genome |
| MWCO | molecular weight cut-off |
| NBT | nitro blue tetrazolium |
| nt | nucleotide |
| (d)NTP | (deoxy)nucleoside triphosphate |
| OD | optical density |
| ORF | open reading frame |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| pH | potentia hydrogenii, -log $[H^+]$ |
| PMSF | phenylmethyl sulphonyl fluoride |
| poly(dI-dC) | synthetic polymer composed of alternating inosine-cytosine nucleotides |
| RNA | ribonucleic acid |
| rRNA | ribosomal RNA |
| RT | reverse transcriptase |
| SAM | *S*-adenosyl-L-methionine |
| SDS | sodium dodecyl sulfate |
| TAP | tobacco acid pyrophosphatase |
| Tris | tris (hydroxymethyl)-aminomethane |
| tRNA | transfer RNA |
| dTTP | deoxythymidin triphosphate |
| Trx | thioredoxin |
| U | unit |
| UTP | uridine triphosphate |
| UV | ultra violet |
| v/v | volume per total volume |
| w/v | weight per volume (g/100ml) |
| X-Gal | 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside |

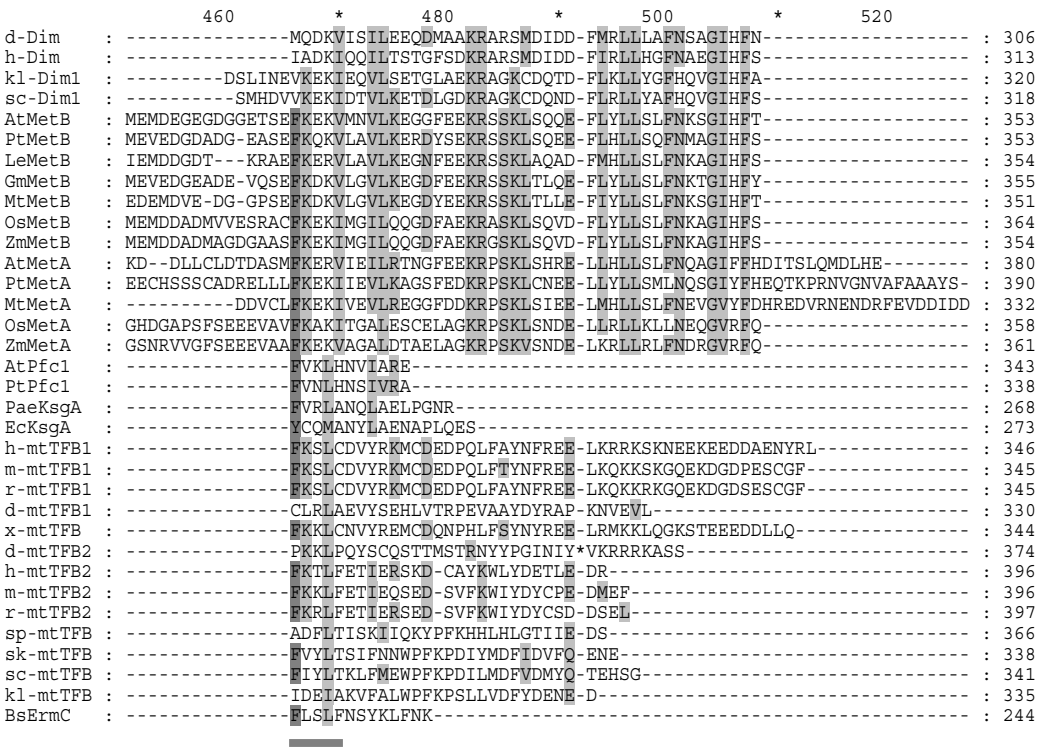
## ANNEX A: Amino acid sequence alignment of methyltransferase-like proteins

Comparison of predicted amino acid sequences of rRNA dimethylases and related mitochondrial transcription factors. Shaded positions are conserved in 40 % (light grey), 70 % (grey) or 100 % (black) of aligned sequences. Regions corresponding to eight conserved sequence motifs of SAM-dependent methyltransferases that are involved in SAM binding (motifs I-III) and in RNA binding and catalysis (motifs IV-VIII) are indicated. Grey bars below mark alignment sections that were included to reconstruct methyltransferase phylogeny. Sequences were aligned of AtMetA (At5g66360), AtMetB (At2g47420) and Pfc1 (At1g01860) from *Arabidopsis thaliana*, PtMetA (fgenesh1_pg.C_LG_XVIII000509), PtMetB (eugene3.00140665) and PtPfc1 (eugene3.00140230) from *Populus trichocarpa*, MtMetA (AC136451.9) and MtMetB (CA922412 and BF632918) from *Medicago truncatula*, GmMetB (CX703076.1) from *Glycine max*., LeMetB (BT013951.1) from *Lycopersicon esculentum*, OsMetA (XP_478337.1B) and OsMetB (AK099711.1) from *Oryza sativa*, ZmMetA (AY108752.1) and ZmMetB (CO520188.1 and CG266357.1) from *Zea mays*, sc-mtTFB (NP_013955.1) and sc-Dim1 (NP_015057.1) from *Saccharomyces cerevisiae*, kl-mtTFB (P87250) and kl-Dim1 (XP_453098.1) from *Kluyveromyces lactis*, sk-mtTFB (P87292) from *Saccharomyces kluyveri*, sp-mtTFB (CAB65608.1) from *Schizosaccharomyces pombe*, h-mtTFB1 (NP_057104.1), h-mtTFB2 (NP_071761.1) and h-Dim (AAH10874.1) from *Homo sapiens*, d-mtTFB1 (CG7319-PA), d-mtTFB2 (CG3910-PA) and d-Dim (NP_651660.1) from *Drosophila melanogaster*, m-mtTFB1 (NP_666186.1) and m-mtTFB2 (NP_032275.1) from *Mus musculus*, r-mtTFB1 (NP_852139.2) and r-mtTFB2 (NP_001008294.1) from *Rattus norvegicus*, x-mtTFB (AAH56010.1) from *Xenopus laevis*, EcKsgA (P06992) from *Escherichia coli*, PaeKsgA (H83571) from *Pseudomonas aeruginosa* and BcErmC (P13956) from *Bacillus subtilis*. Amino acid sequences were derived from genomic and cDNA sequences using the FEX tool at www.softberry.com and the Translate tool at www.expasy.ch.




I
II
III
100 * 120 * 140 * 160 * 180
d-Dim : EKAALR------ATDVVLEIGPGTGNMTVRMLERA---KKVIACEIDTRLAAELQKRV-QATPLQPKLQVLIGDFLKAELPFF------- : 115
h-Dim : DKAALR------PTDVVLEVGPGTGNMTVKLLEKA---KKVVACELDPRLVAELHKRV-QGTPVASKLQVLVGDVLKTDLPFF------- : 122
kl-Dim1 : DKAQIK------PSDIVLEIGPGTGNLTVRILEQA---RKVVAVEFDPRMAAELTKRV-HGTPVEKKLEILLGDFMKTELPYF------- : 123
sc-Dim1 : DKAQIR------PSDVVLEVGPGTGNLTVRILEQA---KNVVAVEMDPRMAAELTKRV-RGTPVEKKLEIMLGDFMKTELPYF------- : 122
AtMetB : QKAGIK------STDVILEIGPGTGNLTKKLLEAG---KEVIAVELDSRMVLELQRRF-QGTPFSNRLKVIQGDVLKTELPRF------- : 120
PtMetB : QKSGIK------STDVVLEIGPGTGNLTKKLLEAG---KMVVAIELDPRMVLELQRRF-QGTAFSNRLKVIQGDVLKTDLPYF------- : 120
LeMetB : QKSGIK------STDVILEIGPGTGNLTKKLLEAG---KSVIAVELDPRMVLELQRRF-QGTPLSNRLKVIQGDVLKCDLPYF------- : 122
GmMetB : EKAGVK------STDILEIGPGTGNLTKKLLEAG---KKVIAVEIDPRMVLELQRRF-QGTPHSNRLTVIQGDVLKTELPYF------- : 120
MtMetB : QKSGIK------TTDVVLEIGPGTGNLTKKLLEAG---KKVIAVEIDPRMVLELNKRF-QGTP-SSRLTVIQGDVLKTELPYF------- : 117
OsMetB : EKAGLK------PTDTVLEIGPGTGNLTKRLLQAGV--KAVVAVELDPRMVLELNRRF-QGDPLASRLKVIQGDVLKCDLPYF------- : 127
ZmMetB : AKAGLK------PTDTVLEIGPGTGNLTKRLLEAGV--KAVVAVELDPRMVLELNRRF-QGHPLSSRLKVIQGDVLKCDLPYF------- : 118
AtMetA : RSSDIR------PTDTVLEIGPGTGNLTMKLLEAA---QNVVAVELDKRMVEILRKRV-SDHGFADKLTIIQKDVLKTDFPHF------- : 155
PtMetA : RKSSIN------PTDTVLEIGPGTGNLTLKLLDVA---SKVVAVEIDKRMVGVLNKRV-KEHGFEDKLSVIREDALKAEFPKF------- : 152
MtMetA : QKSAIN------PSDTVLEIGPGTGNLTLKLLEAS---REVVAIEIDQRMVNILENRA-LKRGLRNKLQVISKDALRTEFPPF------- : 124
OsMetA : RRAALR------PGDAVLEVGPGTGNLTVRLLESPA--ARVSAVEIDPRMVDAVTARV-DALGLAHKLTVIRADAVEAEFPEF------- : 130
ZmMetA : RHAAIS------PGDAVLEVGPGTGNLTARLLASQA--ARVTAVEIDPRMVEAVTARA-AALGLAHKFKVIAGDAVEVEFPEF------- : 141
AtPfc1 : SAADVK------EGDFVLEIGPGTGSLTNVLINLG---ATVLAIEKDPHMVDLVSERF----AGSDKFKVLQEDFVKCHIRSH------- : 160
PtPfc1 : ASANVE------EGDLILEIGPGTGSLTNVLIDAG---ATVLAIEKDAHMAALVRERF----ADTNRFKVLQEDFIKCHIRSH------- : 155
PaeKsgA : RAIHAR------EGQRLLEIGPGQGALTEGLLGSG---ARLDVIELDQDLIPLLKLKF----GLESRFSLHQGDALKFDFASL------- : 98
EcKsgA : SAINPQ------KGQAMVEIGPGLAALTEPVGERL---DQLTVIELDRDLAARLQTHP----FLGPKLTIYQQDAMTFNFGEL------- : 100
h-mtTFB1 : RKAGNL------TNAYVYEVGPGPGGLTRSILNADV--AELLVVEKDTRFIPGLQML---SDAAPGKLRIVHGDVLTFKVEKA-FSESLK : 126
m-mtTFB1 : RKAGSL------ADVYVYEVGPGPGGLTRSILNANV--AELLVVEKDTRFIPGLQML---SDAAPGKLRIVHGDVLTYKIEKA-FPGNIR : 126
r-mtTFB1 : RKAGSL------ADVYVYEVGPGPGGLTRSILNADI--AELLVVEKDTRFIPGLQML---SDAAPGKLRIVHGDVLTYKIEKA-FPDNIR : 126
d-mtTFB1 : KSAGRI-----DPRDLVLEVGPGPGGLTRSILRRHP--QRLLLVEKDPRFGETLQLLKECASPLNIQFDIHYDDILRFNIEQH-IPDTSQ : 133
x-mtTFB : RRAGNL------QNAYVCEVGPGPGGLTRSILNAGV--EELLVVEKDTRFIPGLKML---NEASGGKVRTVHGDILTYRMDRA-FPKHLI : 126
d-mtTFB2 : YLEPHF---QSSGCDTVMELNSGAGYFTRHLLDRESQFRRILLLESMDHFMPKIQEL---HTLYPERVKVRQGDFVNL--WKLV---YMD : 136
h-mtTFB2 : QIYLGK---PSRPPHLLLECNPGPGILTQALLEAG---AKVVALESDKTFIPHLESL---GKNLDGKLRVIHCDFFKLD-PRSG---GVI : 162
m-mtTFB2 : RDLLEH---QNPSRQIILECNPGPGILTGALLKAG---ARVVAFESEKTFIPHLEPL---QRNMDGELQVVHCDFFKMD-PRYQ---EVV : 161
r-mtTFB2 : RDLLEH---QNPSHQLILECNPGPGILTGALLKAG---ARVVAFESEKMFIPHLESL---RKNADGELQVVHCDFFKID-PRYQ---ELV : 162
sp-mtTFB : VKSNLLKEYNS-EKMTILEMAPGPGVTTTSLFNYFQP-KSHVVLESREVFSKPLQKLCTLSDG---RIKWVHQDGYYWQTYEDVYVSKVL : 133
sk-mtTFB : NKLQLQSTYKM-DELKVLDLYPGPSQHSAIFRNIFNP-KQYVLMDSRPDFVKFIQDNF-AGTS--MEL--YQRDPYEWSSYTD------- : 109
sc-mtTFB : DKLDLTKTYKHPEELKVLDLYPGVGIQSAIFYNKYCP-RQYSLLEKRSSLYKFLNAKF-EGSP--LQI--LKRDPYDWSTYSN------- : 110
kl-mtTFB : DKLNLESYYKS-ESLQILDIYAGPLIQSVILNERLKP-KKHVLLFDRLKFVELYQATL-KDHP--SMVN-YNKNPYKWETFLE------- : 113
BsErmC : TNIRLN------EHDNIFEIGSGKGHFTLELVQR-C--NFVTAIEIDHKLCKTTENK--LVDHDNFQV--LNKDILQFKFPKNQ------ : 94
IV V VI
* 200 * 220 * 240 * 260 *
d-Dim : --------------------------DL-CTANVPYQISSPLIF-KLLL-----HRPLF-RCAVLMF--QREFAERLVAKP-GDKL-YCR : 167
h-Dim : --------------------------DT-CVANLPYQISSPFVF-KLLL-----HRPFF-RCAILMF--QREFALRLVAKP-GDKL-YCR : 174
kl-Dim1 : --------------------------DV-CISNTPYQISSPLVF-KLIN------QPKPP-RVSILMF--QREFAMRLLARP-GDSL-YCR : 175
sc-Dim1 : --------------------------DI-CISNTPYQISSPLVF-KLIN------QPRPP-RVSILMF--QREFALRLLARP-GDSL-YCR : 174
AtMetB : --------------------------DI-CVANIPYQISSPLTF-KLLF-----HPTSF-RCAVIMY--QREFAMRLVAQP-GDNL-YCR : 172
PtMetB : --------------------------DI-CVANIPYQISSPLTF-KLLN------HQPSF-RCAIIMF--QREFAMRLVAQP-GDTL-YCR : 172
LeMetB : --------------------------DI-CVANIPYQISSPLTF-KLLA-----HRPLF-RAAVIMF--QREFAMRLVAQP-GDTL-YCR : 174
GmMetB : --------------------------DI-CVANIPYQISSPLTF-KLLKLLK---HEPAF-RAAIIMF--QREFAMRLVAQP-GDKL-YCR : 175
MtMetB : --------------------------DI-CVANIPYQISSPLTF-KLLK-----HQPAF-RCAIIMF--QREFAMRLVAQP-GDKL-YCR : 169
OsMetB : --------------------------DI-CVANIPYQISSPLTF-KLLS------HRPIF-RCAVIMF--QREFAMRLVAQP-GDSL-YCR : 179
ZmMetB : --------------------------DI-CVANIPYQISSPLTF-KLLS------HRPIF-RCAVIMF--QREFAMRLVAQP-GDSL-YCR : 170
AtMetA : --------------------------DL-VVANIPYNISSPLVA-KLVY------GSNTF-RSATLLL--QKEFSRRLLANP-GDSD-FNR : 207
PtMetA : --------------------------DL-VVANIPYGISSPLVA-KLVY------GANPF-RSATLLL--QKEFARRLLAKP-GDSE-FNR : 204
MtMetA : --------------------------DL-VVANIPYGISSPLII-KLIY------ETTPF-RSATLLL--QKEFARRLLANP-GDSE-FNR : 176
OsMetA : --------------------------DV-CVANIPYGISSPLIA-KLLF------GPYRF-RAATLLL--QKEFARRLVAAP-GDSE-YNR : 182
ZmMetA : --------------------------DV-CVANIPYGISSPLIA-KLLF------GPYRF-RTATLLL--QKEFARRLVAAP-GDGE-YNR : 193
AtPfc1 : ----MLSILETRRLSHPDSALA-----KVVSNLPFNISTDVV--KLLLP-----MGDIF-SKVVLLL--QDEAALRLVEPALRTSE-YRP : 230
PtPfc1 : ----MLSMLESMGSLSEKPRYA-----KVVANIPFNISTDVI--KQLLP-----TGDIF-SEIVLLL--QDETALRLVESSLRTSE-YRP : 225
PaeKsgA : ----VES-GEKL---------------RVVGNLPYNISTPLIF-HLLE-----HAPVI-EDMHFML--QKEVVERLAATP-GGGD-WGR : 156
EcKsgA : ----AEKMGQPL---------------RVFGNLPYNISTPLMF-HLFS-----YTDAI-ADMHFML--QKEVVNRLVAGP-NSKA-YGR : 159
h-mtTFB1 : RP----WEDDP-P------------NVHIIGNLPFSVSTPLII-KWLENISCRDGPFVYGRTQMTLTF-QKEVAERLAANTGSKQ--RSR : 195
m-mtTFB1 : RQ----WEDDP-P------------NVHIIGNLPFSVSTPLII-KWLENISLKDGPFVYGRTKMTLTF-QKEVAERLVATTGSKQ--HSR : 195
r-mtTFB1 : RQ----WEDDP-P------------NVHIIGNLPFSVSTPLII-KWLENISLKNGPFVYGRTKMTLTF-QKEVAERLVATTGSKQ--RSR : 195
d-mtTFB1 : R-------------------------IHLIGNLPFAISTRLLI-NWLDDLAARRGAFRRIDTCMTLTF-QQEVAERLCAPVGGEQ--RCR : 194
x-mtTFB : KS----WDDEP-P------------NVHIIGNLPFSVSTPLII-KWLEQVADRTGPFTYGRTQMTLTF-QQEVAERLTASTKNKQ--RSR : 195
d-mtTFB2 : K-MDGGSRVADLLSDVPQKAFT---DDINMLVFGAVGSYPFFK-HLINSLIFQTSLFNLGRCEMILAMPPPIY-IHRTCNNEIGYLIYRS : 220
h-mtTFB2 : KPPAMSSRGLFKNLGIEAVPWTADIPLKVVGMFPSRGEKRALW-KLAYDLYSCTSIYKFGRIEVNMFIGEKEF-QKLMADPGNPDL-YHV : 249
m-mtTFB2 : RPD-VSSQAIFQNLGIKAVPWSAGVPIKVFGILPYKHERRILW-KILFDLYSCESIYRYGRVELNMFVSEKEF-RKLIATPKRPDL-YQV : 247
r-mtTFB2 : RPD-VNSHTIFQNLGIKAVPWSAGVPIKVFGILPNKHERRLLW-KILFDLYSCESIYRYGRVELNMFISEKEF-RKLIATPKRPDL-YQV : 248
sp-mtTFB : DPRIQT-EEEQKLSP--------HRELLFFAHLPHGYAGLLFVSQILDFLSARDWLGIFGRVRVLLWL-PCSPTVTLLGSRGFSK--RSK : 211
sk-mtTFB : -----MIEKEKRFVPNRQSRDKIHNQFLVMANLTGMLGEGLFM-QWLSCIGNKNWLQRFGRVKMLVWV-PEATAHKVLARPGSLI--RAK : 190
sc-mtTFB : -----LIDEERIFVPEVQSSDHINDKFLTVANVTGEGSEGLIM-QWLSCIGNKNWLYRFGKVKMLLWM-PSTTARKLLARPGMHS--RSK : 191
kl-mtTFB : -----MTNEDRVLTPSMQKRDHIHNEFLIAANLTNKKGEQLYV-QYLQCIANQNWMQRFGLVKMLVWI-PQQTARKLFAPFSNKD--RNR : 194
BsErmC : --------------------------SYKIFGNIPYNISTDIIR-KIVFDSI---------ADEIYLIV-EYGFAKRLLNT-------KRS : 141


VII
VIII
280 * 300 * 320 * 340 * 360
d-Dim : LSINTQLLARVDMLMKV----------GKNNFRPPP----------KVESSVVRLEPKN--PP : 208
h-Dim : LSINTQLLARVDHLMKV----------GKNNFRPPP----------KVESSVVRIEPKN--PP : 215
kl-Dim1 : LSANVQMWANVTHIMKV----------GKNNFRPPP----------KVESSVVRIEIKN--PR : 216
sc-Dim1 : LSANVQMWANVTHIMKV----------GKNNFRPPP----------QVESSVVRLEIKN--PR : 215
AtMetB : LSVNTQLYARVSHLLKV----------GKNNFRPPP----------KVDSSVVRIEPRR--PG : 213
PtMetB : LSVNTQLYARVSHLLKV----------GKNNFRPPP----------KVDSSVVRIEPRK--PR : 213
LeMetB : LSVNTQLLARVSHLLKV----------GKNNFRPPP----------KVDSSVVRIEPRG--PL : 215
GmMetB : LTVNTQLHARVFHLLKV----------GRNNFRPPP----------KVDSSVVRIEPRK--PR : 216
MtMetB : LTVNTQLHARISHLLKV----------GRNNFRPPP----------KVDSSVVRIEPKK--PR : 210
OsMetB : LSVNVQLLSRVSHLLKV----------GRNNFRPPP----------KVDSSVVRIEPRK--PL : 220
ZmMetB : LSVNVQLLSRVSHLLKV----------GRNNFRPPP----------KVDSSVVRIEPRK--PL : 211
AtMetA : LAVNVKLVADVKFVMDV----------SKREFVPPP----------KVDSSVIRITPKE--II : 248
PtMetA : LAVNVKLVADVEFVMNV----------SKRDFFPVP----------KVDSSVVIIRPKD--RI : 245
MtMetA : LAVNIKLLADVEFVMDV----------SKRDFLPSP----------KVDSSVVIIRPKV--NV : 217
OsMetA : LAANVRMVADARLLMDV----------SKRDFVPMP----------KVDSSLVEIRPRA--AE : 223
ZmMetA : LAANVRLVADVRLLMDV----------SKRDFVPMP----------RVDSSLVEIRPRG--TV : 234
AtPfc1 : INILINFYSEPEYNFRV----------PRENFFPQP----------KVDAAVVTFKLKHPRDY : 273
PtPfc1 : INIFVNFYSDPEYKFKV----------PRSNFFPQP----------KVDAAVVRFKLKQAVDY : 268
PaeKsgA : LSIMVQYHCRVEHLFNV----------GPGAFNPPP----------KVDSAIVRITPFAEPPH : 199
EcKsgA : LSVMAQYYCNVIPVLEV----------PPSAFTPPP----------KVDSAVVRLVPHATMPH : 202
h-mtTFB1 : LSVMAQYLCNVRHIFTI----------PGQAFVPKP----------EVDVGVVHFTPLI-QPK : 237
m-mtTFB1 : LSIMAQYLCNVEHLFTI----------PGKAFVPKP----------KVDVGVVHLTPLI-EPK : 237
r-mtTFB1 : LSIMAQYLCNVEHLFTI----------PGKAFVPKP----------EVDVGVVHLMPLV-QPK : 237
d-mtTFB1 : LSVMSQVWTEPVMKFTI----------PGKAFVPKP----------QVDVGVVKLIPLK-RPK : 236
x-mtTFB : LSIMSQYLCNVKNCFTI----------PGRAFIPKP----------KVDVGVVHLTPFV-QPK : 237
d-mtTFB2 : TSVLFQILFEHKFIAKV----------PREDFLPQQMAYSPTKSSKLGKVQSINPEYLYLVKFTPRRNLHE : 281
h-mtTFB2 : LSVIWQLACEIKVLHME----------PWSSFDIYTRK-GPLENPKRRELLDQLQQKLYLIQMIPRQNLFT : 309
m-mtTFB2 : MAVLWQVACDVKFLHME----------PWSSFSVHMEN-GHLEKSKHGESVNLLKQNLYLVRMTPRRTLFT : 307
r-mtTFB2 : LGVLWQVACEIKFLHME----------PWSSFSVHAEN-GHLEKSKHSESLNLLKQNLYLVRMTPRRTLFT : 308
sp-mtTFB : TSVFREAFTDSRVLAASE-STLQKLCMGYSKEAKENYQISPNPLLVSPTPITSEPH----------KEQLTLVEMCSKPQDK- : 282
sk-mtTFB : CSVVTEAFTDTKLVATSDSSTLQK----FSSSLLEGH----DPIIFSTRDTWLNS----------GKPISLLEVNPIDHDI- : 253
sc-mtTFB : CSVVREAFTDTKLIAISDANELKG----FDSQCIEEW----DPILFSAAEIWPTK----------GKPIALVEMDPIDFDF- : 254
kl-mtTFB : LTLLSELATNTKLVATSE-NSVKK----FLPDCIEKF----DPVIIPSDN---KS----------PDDLSLVEINPRDHSI- : 253
BsErmC : LALFLMAEVDISILSMV----------PREYFHPKP----------KVNSSLIRLNRKKSRIS : 184
* 380 * 400 * 420 * 440 *
d-Dim : P-PVNFTEWDGLT--RIAFLRKNKTLAATFKVTSVLEMLEKNYKLYRSLRNEPIEDDFK---------- : 264
h-Dim : P-PINFQEWDGLV--RITFVRKNKTLSAAFKSSAVQQLLEKNYRIHCSVHNIIIPEDFS---------- : 271
kl-Dim1 : P-QVDFNEWDGLL--RIVFVRKNRTIAAGFKSTTVLEILEKNYKAFLATQSAVPTTSSG---------- : 272
sc-Dim1 : P-QVDYNEWDGLL--RIVFVRKNRTISAGFKSTTVMDILEKNYKTFLAMNNEMVDDTKG---------- : 271
AtMetB : P-QVNKKEWDGFL--RVCFIRKNKTLGSIFKQKSVLSMLEKNFKTLQAVLASLQNNGEPALNTTS--MDLGDQSMGMEDDDNE-M-DDDD : 296
PtMetB : P-QVNPKEWDGFI--RICFIRKNKTLGSIFRIKNVLSMLEKNYKTLQA-LQQLQNGSSGSTNAEMDILGLGDSKEDHSMDMDD-G-TDDE : 297
LeMetB : T-PVNFKEWDGLV--RICFNRKNKTIGSIFRQKSVLNILEKNYRTLQA-LQFSEKAPSNDMEMALDVSTLGESFGDLSMDADD-GNDDDD : 300
GmMetB : I-EVKQKEWDGFL--RICFNRKNKTLGSIFRQKSVISLLEKNYRTVRA-LELGQEDSLKEVDAKMDFSSFGD---DRGMEMDDDGADDDE : 299
MtMetB : H-EVNQKEWDGFL--RICFNRKNKTLGAIFRQKNVISMLEKNYKTVQA-LKLSQEGLLKEADTKVDFSNFADFVDDQGMEMDDDGVDDND : 296
OsMetB : P-PVSFKEWDGLV--RLCFNRKNKTLGAIFKQKRVLELLEKNYKTMQSLQLTSDAEKGEEKMSPDDVALLSSMVDDMNMESSYENDDDDE : 307
ZmMetB : P-PVSFKEWDGLV--RICFNRKNKTLGSLFKQKRVLELLEKNYKTMQSLQTVQDAEMGEEKMSADDVALLANMVEDMSMETGYDKEDD-E : 297
AtMetA : P-DVNVQEWLAFT--RTCFGKKNKTLGSMFRQKKKVMELQSLSAGRHGSNVE----------VMNQTGGDSDSDVEEDG : 314
PtMetA : P-DVNLDEWWAFT--KNCFGKKNKTLGATFKQKKKVIELFRLSKMTSSNGEE----------INRNQYVCADDDGDNEEESDG : 315
MtMetA : P-AVNLHQWRAFT--RTCFNNKNKTLGATFKNKRKVLELLEFNNVSGLVGEQ---------- : 266
OsMetA : P-NVDLAEWLAFT--RSCFGQKNKTLGAIFKQKRKVLELFRRSRCAEERCDGNANAGGGS---------RLIALGGGDDDGDMSDGDSNE : 301
ZmMetA : P-GVDVSEWLAFT--RVCFGQKNKTLAAIFKQKRMVVELLRRS-LRTERCAGGVSLGPLE----------VDDSGEDDCGGNDD : 304
AtPfc1 : PDVSSTKNFFSLV--NSAFNGKRKMLRKSLQHISSSPDIEKALGVAGLPATSRPEELTLDD---------- : 332
PtPfc1 : PVVSSTKSFFSMV--NSAFNGKRKMLRKSLQHICTPVEIEEALQNVGCLATSRPEELTLDD---------- : 327
PaeKsgA : P-ARDPKLLERVV--REAFNQRRKTLRNTLKPLLSVEDIEAA----EVDPTLRPEQLDLAA---------- : 253
EcKsgA : P-VKDVRVLSRIT--TEAFNQRRKTIRNSLGNLFSVEVLTGM----GIDPAMRAENISVAQ---------- : 256
h-mtTFB1 : IEQ-PFKLVEKVV--QNVFQFRRKYCHRGLRMLFPEAQRLESTGRLLELADIDPTLRPRQLSISH---------- : 299
m-mtTFB1 : IKQ-PFKLVEKVV--QNAFQFRRKYCHRGLGMLFPEAQRLESTGRLLQLADIDPTLRPTHLSLMH---------- : 299
r-mtTFB1 : IKQ-PFKLVEKVV--QNVFQFRRKYCHRGLGMLFPEAQRLESTGRLLQLADIDPTLRPTHLSLMH---------- : 299
d-mtTFB1 : TQL-PFHLVERVV--RHIFSMRQKYCRRGYGTLLPPEDREEVAEKLFQRAEVQDTLRPFELTVEQ---------- : 298
x-mtTFB : IEQ-PFKLVEKVV--RCIFQFRRKYCHHGVSILFPEEIRIQLTEQMLRLADVDPTLRPTELTMTH---------- : 299
d-mtTFB2 : LCQSQDLPALWFFIKQNYVSRRNRIIPNLEKWVPGCGPRLI---INPKSSESVTPIYPDEL---------- : 339
h-mtTFB2 : KNLTPMNYNIFFHLLKHCFGRRSATVIDHLRSLTPLDARDILMQIGKQEDEKVVNMHPQD---------- : 369
m-mtTFB2 : ENLSPLNYDIFFHLVKHCFGKRNAPIIRHLRSLSTVDPINILRQIRKNPGDTAARMYPHD---------- : 367
r-mtTFB2 : ENLSPLNYDMFFHLVKHCFGKRNAPIIHHLRSLSTVDPINILRQIRKRPGDTAAKMYPHD---------- : 368
sp-mtTFB : --QLSIPVFESIV--RILLTCKATSLSKSIYYLGPGAETLLPSFTQCGINIDMPVGLLSA---------- : 338
sk-mtTFB : --DLDNWDYVTKH--LLILKSTPLHTAIDSLGHGGKQYFSEKVEDKLLMDKCPKDLTNKE---------- : 309
sc-mtTFB : --DVDNWDYVTRH--LMILKRTPLNTVMDSLGHGGQQYFNSRITDKDLLKKCPIDLTNDE---------- : 310
kl-mtTFB : --DLDHWDFVTQK--LMILKSKPVEEMIEILGHGARDWFISRL-DPVLLKKKPYELTYLE---------- : 308
BsErmC : HKDKQKYNYFVMK--WVNKEYKKIFT---------KNQFNNSLKH--AGIDDLNNISFEQ---------- : 231



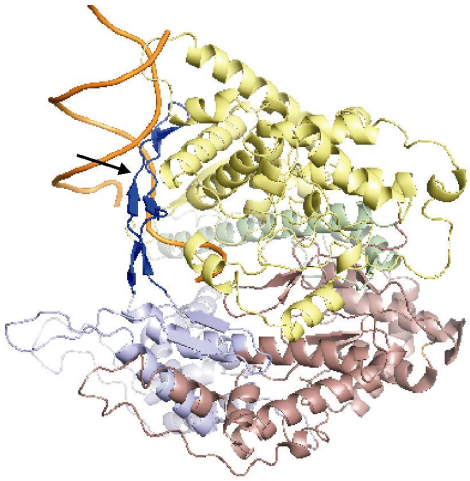

## ANNEX B: Predicted subcellular targeting of plant methyltransferase-like proteins

Predictions of targeting to the mitochondrion (M), nucleus (N), or plastid (P) by the computer algorithms listed in Table 11 are documented; Mitoprot predictions are given as probabilities of mitochondrial targeting. Proteins are designated as in Annex A.

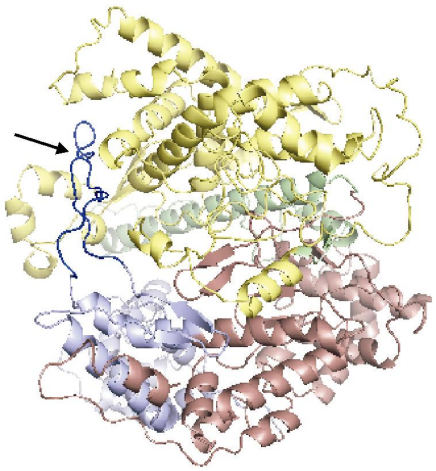
| Protein | Predotar | TargetP | Mitoprot | Psort | iPsort |
|---|---|---|---|---|---|
| AtMetA | Possibly **M** | **M** | **0.59** | **M** or **N** | **M** |
| AtMetB | Not M or P | Not M or P | 0,25 | **N** | Not M or P |
| AtPfc1 | **P** | **P** | 0.24 | Not M or P | **P** |
| PtMetA | Possibly **M** | **M** | **0.99** | **N** | **M** |
| PtMetB | Not M or P | Not M or P | 0.47 | Not M or P | **M** |
| PtPfc1 | Possibly **P** | **P** | **0.79** | **P** or **N** | **P** |
| MtMetA | **M** | **M** | **0.93** | **M** | Not M or P |
| MtMetB | Not M or P | Not M or P | 0.09 | Not M or P | Not M or P |
| GmMetB | Not M or P | Not M or P | 0.23 | Not M or P | Not M or P |
| LeMetB | Not M or P | Not M or P | 0.23 | Not M or P | **M** |
| OsMetA | **M** | **M** | **0.78** | **M** | **M** |
| OsMetB | Not M or P | Not M or P | 0.16 | Not M or P | Not M or P |
| ZmMetB | Not M or P | Not M or P | 0.41 | Not M or P | Not M or P |
| ZmMetA | **M** | **M** | **1.0** | **M** | **M** |

## ANNEX C: Threaded structural models of Arabidopsis RpoTm and RpoTmp

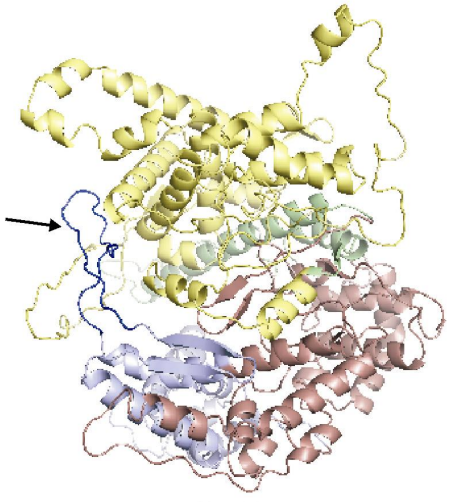
The RpoTm and RpoTmp structural models (right) are based on T7 RNA polymerase crystal structures (Cheetham, et al., 1999; Cheetham and Steitz, 1999). The N-terminal domain of the T7 enzyme is coloured pale yellow; the thumb, palm and fingers subdomains are shown in pale green, pale red and pale blue, respectively; the specificity loop (arrow) is bright blue and the double helix of the DNA template is orange (compare I.3.2.2 and Figure 5). Corresponding regions of the RpoTm and RpoTmp structural models are coloured accordingly. The RpoTm and RpoTmp models exclude the first 104 (RpoTm) and 150 (RpoTmp) amino acids from the N-terminus, which are not conserved with the T7 RNA polymerase. The secondary structure of regions corresponding to the T7 enzyme specificity loop (arrows) is unresolved in the RpoTm and RpoTmp models. The coordinates of the threaded models of RpoTm and RpoTmp were derived from www.bmm.icnet.uk/servers/3djigsaw/ (Bates, et al., 2001; Bates and Sternberg, 1999; Contreras-Moreira and Bates, 2002) with the program choosing the T7 RNA polymerase/promoter DNA complex [PDB number 1cez (Cheetham, et al., 1999)] and the structure of a transcribing T7 RNA polymerase initiation complex [PDB number 1qln (Cheetham and Steitz, 1999)] as the starting structures to create the threaded model coordinates for RpoTm and RpoTmp, respectively. The images were generated using PyMol v0.98 (http://delsci.com/rel/0_98/). Choosing the T7 enzyme/promoter DNA complex (PDB number 1cez) as structure template and providing the RNA polymerase alignment shown in Annex A, SWISS-MODEL [http://swissmodel.expasy.org/, (Guex and Peitsch, 1997; Peitsch, 1995; Schwede, et al., 2003)] was independently used to build 3D models of RpoTm and RpoTmp polypeptide regions corresponding to thumb, palm and fingers subdomains of the T7 enzyme; the final coordinates generated were highly similar (data not shown).



T7



RpoTm



RpoTmp

## ANNEX D: Accession numbers of RNA polymerase sequences

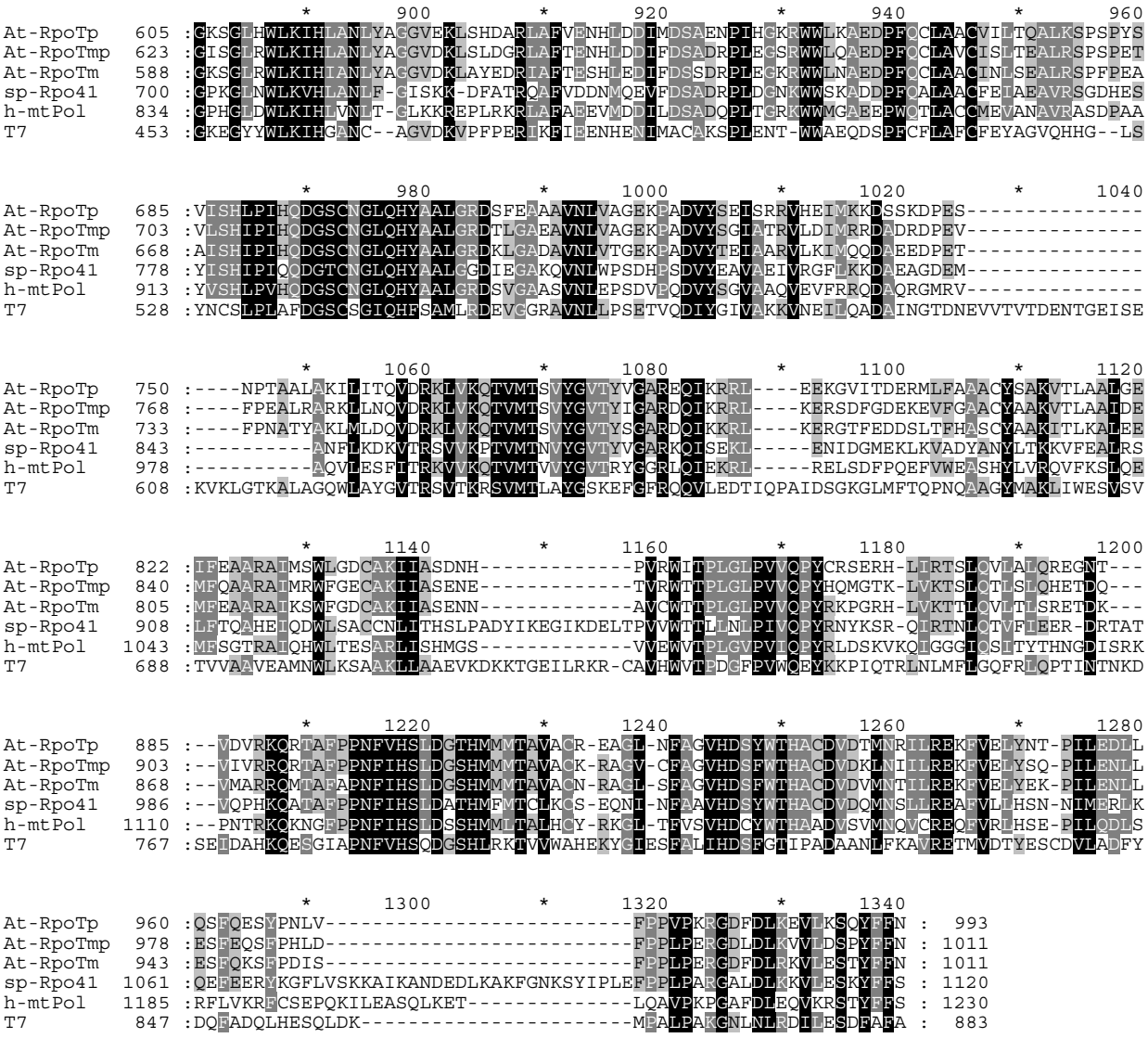
Accession numbers of amino acid sequences of bacteriophage and phage-type RNA polymerases included in the alignment shown in Fig. 32 and Annex E. Protein targeting to the mitochondrion (M) and/or plastid (P) is indicated.

| Protein | Accession number | Source | Targeting |
|---|---|---|---|
| At-RpoTm | CAA69331 | *Arabidopsis thaliana* | M |
| At-RpoTmp | CAC17120 | *Arabidopsis thaliana* | M, P |
| At-RpoTp | CAA70210 | *Arabidopsis thaliana* | P |
| Ns-RpoTm | CAC95019 | *Nicotiana sylvestris* | M |
| Ns-RpoTmp | Q8VWF8 | *Nicotiana sylvestris* | M, P |
| Ns-RpoTp | P69242 | *Nicotiana sylvestris* | P |
| Ta-RpoTm | AAF3249 | *Triticum aestivum* | M |
| Ta-RpoTp | AAB01085 | *Triticum aestivum* | P |
| Zm-RpoTm | AAD22976 | *Zea mays* | M |
| Zm-RpoTp | AAD22977 | *Zea mays* | P |
| sc-Rpo41 | P13433 | *Saccharomyces cerevisiae* | M |
| sp-Rpo41 | CAB16197 | *Schizosaccharomyces pombe* | M |
| h-mtPol | AAC06147 | *Homo sapiens* | M |
| T7 | NP_041960 | Bacteriophage T7 | - |
| T3 | NP_523301 | Bacteriophage T3 | - |
| phiYeO3-12 | NP_052071 | Bacteriophage phiYeO3-12 | - |
| SP6 | P06221 | Bacteriophage SP6 | - |

## ANNEX E: T7 phage and phage-type RNA polymerase sequence alignment

Comparison of amino acid sequences of RpoTm, RpoTmp and RpoTp from *Arabidopsis thaliana*, Rpo41 from *Schizosaccharomyces pombe*, the mitochondrial RNA polymerase from *Homo sapiens*, and the T7 bacteriophage RNA polymerase (see Annex D for accession numbers). Shaded positions are conserved in 60 % (light grey), 80 % (grey) or 100 % (black) of aligned sequences. N-terminal portions of organellar enzymes, which do not align to the T7 sequence, are not displayed. Amino acid sequences were compared using the Multalin algorithm (Corpet, 1988); the alignment was refined according to a secondary structure-based alignment of RpoTm to the T7 sequence generated using the GenTHREADER tool at http://bioinf.cs.ucl.ac.uk/psipred/ (Jones, 1999; McGuffin, et al., 2000; McGuffin and Jones, 2003).

```
                    *       340         *       360         *       380         *       400
At-RpoTp     94 :SYAGTIKGASLIEELENPVERNGLSGRRRLFMQDPPWISALFLKG---LSKMVDQTLKIERKD-----------------
At-RpoTmp    92 :KGFRCLSGITRREEFSKSERCLSGTLARGYTSVAEEEVLSTDVEEEPEVDELLKEMKKEKKRE-------SHRSWRMKKQ
At-RpoTm     51 :NPVNEMGGLSSFRHGQCYVFEGYATAAQAIDSTDPEDESSGSDEVNELITEMEKETERIRKKA--------RLAAIPPKR
sp-Rpo41    179 :-------ITNSTIVGESENGKDQNGDSSLKEKQPDVETTVTKSANLNALRSSLSSLLTESIDLPIDEVSLEFGNQGDTFN
h-mtPol     312 :DAGT---IERCLEQMSQEGLKLQALFTAVLLSEEDRATVLKAVHKVKPTFSLPPQLPPPVNTSKLL-----RDVYAKDGR
T7            1 :-------------------------------------------------MNTI-NIAKND----FSDIELAAIPFNT

                    *       420         *       440         *       460         *       480
At-RpoTp     154 :----IDKRKFDSLRRRQVKEETEAWERMVDEYRDLEKEMCEKNLAPNLPYVKHMFLGWFQPLKDVI-EREQKLQKNKSKK
At-RpoTmp    165 :DQFGMGRTKFQNLWRRQVKIETEEWERAAAEYMELLTDMCEQKLAPNLPYVKSLFLGWFEPLRDAI-AKDQELYRLGKS-
At-RpoTm     123 :VIAGMGAQKFYMLKQRQVKMETEEWERAARECREILADMCEQKLAPNLPYMKSLFLGWFEPVRNAI-QDDLDTFKIKKG-
sp-Rpo41     252 :LARQKLLEKSAILSAAEVWKSEHESV-LNRGNLQVPKNVSSLFYSWYV-QLEQLFKEEI-SLIDDL-ALNESLDKKND--
h-mtPol      384 :VSYPKLHLPLKTLQCFFEKQLHMELASRVCVVSVEKPTLPSKEVKHARKTLKTLRDQWEKALCRALRETKNRLEREVYEG
T7            24 :----LADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAGEVADNAAAKPLITTLLPKMIARI-NDWFEEVKAKRGK

                    *       500         *       520         *       540         *       560
At-RpoTp     229 :VRAAYAPHIELLPADKMAVIVMHKMMGLVMSGHED-GCIQVVQAAVSIGIAIEQEVRIHNFLKRTRKNNA----------
At-RpoTmp    243 :-KATYAHYLDQLPADKISVITMHKLMGHLMTGGDN-GCVKVVHAACTVGDAIEQEIRICTFLDKKKKGDD-----NEESG
At-RpoTm     201 :-KIPYAPFMEQLPADKMAVITMHKMMGLLMTNAEGVGIVKLVNAATQIGEAVEQEVRINSFLQKKNKKNATDKTINTEAE
sp-Rpo41     326 :-RLIYGPFLKLLSSKKLAALIIMEVAQLSTNPRYD-RGARVTTLLGGLGRSFEREF-LSEQIQRQEKNKSY---------
h-mtPol      464 :-RFSLYPFICLLDERERVVRMLIQVLQALPAQGESFTTLARELSARTFSRHVVQRQ-RVSGQVQALQNHY------RKYLC
T7            99 :-RPTAFQELQEIKPEAVAYITIKTTLACLTSADN----TTVQAVASAIGRAIEDEARFGRIRDLEAKHF---------KK

                    *       580         *       600         *       620         *       640
At-RpoTp     298 :-GDSQEELKEKQLLRKRVNSLIRRKRIIDALKVVKS-EGTKPWGRATQAKLGSRLLELLIEAAYVQPPLTQSGDSIP--E
At-RpoTmp    316 :GVENETSMKEQDKLRKKVNELIKKQKLSAVRKILQSHDYTKPWIADVRAKVGSRLIELLVRTAYIQSPADQQDNDLP--D
At-RpoTm     280 :NVSEEIVAKETEKARKQVTVLMEKNKLRQVKALVRKHDSFKPWGQEAQVKVGARLIQLLMENAYIQPPAEQFDDGPP--D
sp-Rpo41     394 :-KDKKRLKELFNDPRKFRQAVKNLRLSNTRDNIVLN-PSVDSWPSAIVMKVGSVALCLLLSVAKIEVTAKDLSTGGILKQ
h-mtPol      536 :LLASDAEVPEPCLPRQYWEELGAPEALR----------EQPWPLPVQMELGKLLAEMLVQATQMPCSLDKPHRSSRL--
T7           165 :NVEEQLNKRVGHVYKKAFMQVVEADMLS---KGLLGGEAWSSWHKEDSIHVGVRCIEMLIESTGMVSLHRQNAG------

                    *       660         *       680         *       700         *       720
At-RpoTp     374 :FRPAFRHRFKTVTKYPGSKLVRRYGVIECDSLLLAGLDKSA---KHMLIPYVPMLVPFKRWKGYDKCGYLFLPSYIMRTH
At-RpoTmp    394 :VRPAFVHTFKVAK--GSMNSGRKYGVIECDPLVRKGLEKSG---RYAVMPYMPMLVPPLKWSGYDKCAMLFLTSYIMKTH
At-RpoTm     358 :IRPAFKQNFRTVTL-ENTKTSRRYGCIECDPLVLKGLDKSA---RHMVIPYLPMLIPPQNWTGYDQCAHFFLPSYVMRTH
sp-Rpo41     472 :EVAAFVHTY-------QYSNGRKVGMIVPHVEFYKLLSRDIEK-PHLHPQLLPMLVTPKPWTSWIDCGYYYSRQPLVRLK
h-mtPol      603 :-VPVLYHVY-------SFRNVQQIGILKPHPAYVQLLEKAAEPTLTFEAVDVPMLCPPLPWTSPHSCAELLSPTKLMRTV
T7           236 :------------------VVGQDSETIELAPEYAEAIATRAGALAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALV

                    *       740         *       760         *       780         *       800
At-RpoTp     451 :GSKKQQDALKDISHKT-AHRVFEALDTLGNTKWRVNRNILDVVERLWA--DGGNIAGLVN---REDVPIPEKPSSEDPEE
At-RpoTmp    469 :GAKQQREALKSAPKGQ-LQPVFEALDTLGSTKWRVNKRVLTVVDRIWS--SGGCVADMVD---RSDVPLPEKPDTEDEGI
At-RpoTm     434 :GAKQQRTVMKRTPKEQ-LEPVYEALDTLGNTKWKINKKVLSLVDRIWA--NGGRIGGLVD---REDVPIPEEPERDQEK
sp-Rpo41     544 :GALEQVDYLMKASENGQLDELFKAVSSLCKVSWRINQRLFNVLIRIW---NSGEKFLSLPPREVKCDMPPYPKNSINPRD
h-mtPol      675 :EGATQHQELLETCPPTALHGALDALTQLGNCAWRVNGRVLDLVLQLEQAKGCPQLGVPAPPSEAPQPEEAHLPHSAAPAR
T7           298 :RTHSKKALMRYEDVY--MPEVYKAINIAQNTAWKINKKVLAVANVITK--WKHCPVEDIPAIEREELPMKPEDIDMNPEA

                    *       820         *       840         *       860         *       880
At-RpoTp     525 :LQSWKWSARKANKINRERHSLRCDVELKLSVARKMKDEEGFYYPHNLDFRGRAYPMHPHLNHLSSDLCRGTLEFAEGRPL
At-RpoTmp    543 :LKKWKWEVKSAKKVNSERHSQRCDTELKLSVARKMKDEEAFYYPHNMDFRGRAYPMPPHLNHLGSDLCRGVLEFAEGRPM
At-RpoTm     508 :FKNWRWESKKAIKQNNERHSQRCDIELKLEVARKMKDEEGFYYPHNVDFRGRAYPIHPYLNHLGSDLCRGILEFCEGKPL
sp-Rpo41     621 :KVIWHTRRKELAALKTGAHSQRCDFNYKLELARAFLN-EKFYFPHSLDFRGRAYPLSSHLHHVSNDVCRGLLEFSTGKPL
h-mtPol      755 :KAELRRELAHCQKVARFMHSLRAEALYRLSLAQHLRD-RVFWLPHNMDFRGRTYECPPHFNHLGSDVARALLEFAQGRPL
T7           374 :LTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSMFNPQGNDMTKGLLTLAKGKPI
```


At-RpoTp
At-RpoTmp
At-RpoTm
sp-Rpo41
h-mtPol
T7

## Curriculum Vitae

Kristina Kühn
Born: August 27, 1974; Stollberg, Germany

**Employment:**

| | |
|---|---|
| Aug. 2005 - present, | Research Associate, |
| 2002 - 2004 | Institut für Biologie/Genetik, Humboldt-Universität zu Berlin, Germany (Collaborative Research Centre "Molecular Physiology, Energetics, and Regulation of Primary Metabolism in Plants" – SFB429)<br>Adviser: Professor Dr. Thomas Börner |
| Aug. - Dec. 2001 | Research Associate,<br>Institut für Biologie/Genetik, Humboldt-Universität zu Berlin |
| 1998 - 2000 | Student Assistant,<br>Institut für Biologie/Genetik, Humboldt-Universität zu Berlin |

**Education:**

| | |
|---|---|
| 2002 - present | Doctoral studies<br>Humboldt-Universität zu Berlin, Germany<br>Doctoral Thesis: Analysis of Components of the Transcription Machinery in Arabidopsis Mitochondria<br>Thesis supervisor: Professor Dr. Thomas Börner |
| 2001 | Diploma in Biology (Grade 1.0),<br>Humboldt-Universität zu Berlin, Germany<br>Diploma Thesis: Phage-related transcription systems of *Arabidopsis thaliana*: Functional characterization of recombinant components. |
| 1998/1999 | Biochemistry studies,<br>Imperial College of Science, Technology and Medicine, London, UK<br>Research Project: Isolation and characterisation of PSII-LHCII supercomplexes from *Chlamydomonas reinhardtii* |
| 1995 - 2001 | Biology studies,<br>Humboldt-Universität zu Berlin, Germany<br>Advanced courses in Biochemistry, Molecular Biology and Microbiology |
| 1994 - 1995 | Academic Year at the EF London School of English, London, UK<br>Certificate: Cambridge Certificate of Proficiency in English (Grade A) |
| 1990 - 1994 | High school in Berlin |
| 1981 - 1990 | Primary school in Berlin |

**Scholarships/Awards:**

Ph.D. scholarship by the „Berliner Programm zur Förderung der Chancengleichheit für Frauen in Forschung und Lehre“ (January 2005 - July 2005)

Scholarship awarded by the Studienstiftung des deutschen Volkes (February 1998 - June 2001)

Research grant awarded by the Studienstiftung des deutschen Volkes for a short-term research project at Yale University (Chemistry Dept.), New Haven/CT, USA (May 1999)

SOKRATES studentship (October 1998 - March 1999)

Studentship by the Studienstiftung des deutschen Volkes for biochemistry studies at the Imperial College of Science, Technology and Medicine, London, UK (October 1998 - March 1999)

Berlin, 30. November 2005

## Publications and Conference Abstracts

Publications:

Holec, S., Lange, H., **Kühn, K.**, Alioua, M., Börner, T. and Gagliardi, D. (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and PNPase. *Mol Cell Biol* **26**, 2869-2876

**Kühn, K.**, Weihe, A., Börner, T. (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Res* **33**, 337-46

Morais, F., **Kühn, K.**, Stewart, D.H., Barber, J., Brudvig, G.W., Nixon, P.J. (2001) Photosynthetic water oxidation in cytochrome *b*-559 mutants containing a disrupted haem-binding pocket. *J Biol Chem* **276**, 31986-31993

**Kühn, K.**, Bohne, A.-V., Liere, K., Weihe, A., Börner, T. Accurate *in vitro* transcription of mitochondrial genes by the phage-type RNA polymerase RpoTm from Arabidopsis. *Manuscript in preparation.*

Conference presentations:

Kühn, K., Weihe, A., Liere, K., Börner, T. (2005) Components of the transcription machinery in Arabidopsis mitochondria. 7th International Congress on Plant Mitochondria, Obernai, France

Kühn, K., Weihe, A., Börner, T. (2004) Components of plant mitochondrial transcription. 9th International Congress of Endocytobiology and Symbiosis, Jena, Germany

Kühn, K., Hedtke, B., Liere, K., Emanuel, C., Richter, U., Weihe, A., Börner, T. (2002) Plant mitochondrial phage-type RNA polymerases. 6th International Congress on Plant Mitochondria, Perth, Australia

Conference posters:

Kühn, K., Swiatecka, M., Liere, K., Weihe, A., Börner, T. (2004) Mitochondria and plastids: Complex machineries transcribe simple genomes. 15th International Conference on Arabidopsis Research, Berlin, Germany

Kühn, K., Weihe, A., Börner, T. (2004) Mitochondrial genes in Arabidopsis possess multiple promoters. 17. Tagung Molekularbiologie der Pflanzen, Dabringhausen, Germany

Kühn, K., Weihe, A., Liere, K., Börner, T. (2003) Components of mitochondrial transcription in *Arabidopsis thaliana*. 7th International Congress of Plant Molecular Biology, Barcelona, Spain

Kühn, K., Weihe, A., Liere, K., Börner, T. (2003) Components of mitochondrial transcription in *Arabidopsis thaliana*. FEBS Advanced Course: Origin and Evolution of Mitochondria and Chloroplasts, Hvar, Croatia

Kühn, K., Liere, K., Weihe, A., Börner, T. (2002) Phage-related transcription systems of *Arabidopsis thaliana*: Functional characterisation of recombinant components. Botanikertagung, Freiburg i. Br., Germany

Berlin, 30. November 2005

**Danksagung**

Meinem akademischen Lehrer Herrn Professor Thomas Börner gilt mein herzlicher Dank für die Möglichkeit, an seinem Lehrstuhl ein überaus interessantes Thema zu bearbeiten und meine wissenschaftlichen Arbeiten selbständig zu vertreten. Seine grosszügige Unterstützung und wertvollen Anregungen haben wesentlich zum Entstehen dieser Arbeit beigetragen.

Für die finanzielle Förderung der Arbeit möchte ich der Deutschen Forschungsgemeinschaft und dem „Berliner Programm zur Förderung der Chancengleichheit für Frauen in Forschung und Lehre“ danken.

Meinem Projektleiter Andreas Weihe und Karsten Liere danke ich für vielfältige praktische Ratschläge und gedanklichen Austausch.

Wolfgang Hess und den Teilnehmern der Tagung ICPMB 2005 in Obernai, besonders Dominique Gagliardi, gilt mein Dank für stimulierende wissenschaftliche Diskussionen.

Für ihr motivierendes Interesse und ihre immerwährende Diskussionsbereitschaft bin ich Elke Dittmann überaus dankbar, ebenso Uwe Richter. Ihm und Holger Jenke-Kodama danke ich ausserdem für eine Einführung in die molekulare Evolution. Christian Schmitz-Linneweber sei gedankt für seine wertvollen Anmerkungen zu dieser Arbeit.

Conny Stock und Martin Meixner gilt mein Dank für ihre praktische Unterstützung.

Allen Kollegen und ehemaligen Mitarbeitern der Arbeitsgruppen Genetik und Molekulare Ökologie danke ich für ein herzliches und stimulierendes Arbeitsumfeld.

Für ihre unverzichtbare persönliche Unterstützung gilt mein besonderer Dank meinen Eltern.

Ein herzliches Dankeschön allen Freunden für Motivation und Geduld.

**Eidesstattliche Erklärung**

Hiermit versichere ich, die vorliegende Dissertation eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben. Diese Arbeit wurde keiner anderen Prüfungsbehörde vorgelegt.

Berlin, 30. November 2005